Essays on Behavioral Economics:

Empirical Studies on Risk, Morality and Framing

Inaugural-Dissertation

zur Erlangung des akademischen Grades eines Doktors

der Wirtschafts- und Sozialwissenschaften

der Wirtschafts- und Sozialwissenschaftlichen Fakultät

der Christian-Albrechts-Universität zu Kiel

vorgelegt von

Diplom-Volkswirt Hauke Jelschen

geboren am 30.11.1983

in Oldenburg (Oldb)

Kiel, 2021

Gedruckt mit Genehmigung der
Wirtschafts- und Sozialwissenschaftlichen Fakultät
der Christian-Albrechts-Universität zu Kiel

Dekan: Prof. Dr. Kai Carstensen
Erstbegutachtung: Prof. Dr. Dr. Ulrich Schmidt
Zweitbegutachtung: Prof. James Konow, PhD

Datum der mündlichen Prüfung: 01.02.2021

Acknowledgements

It has been kind of a long haul and I would not have been able to finish this dissertation without the support, patience and help of many people. I would like to name a few here, but this list is certainly not intended to be exhaustive. First and foremost, I thank my supervisor Ulrich Schmidt. It has not always been smooth sailing, but knowing to have him in my corner meant a lot to me throughout this whole process. My second thanks goes to James Konow, who hired me to work at the Chair of Economics and Ethics and offered plenty of opportunities to familiarize myself with the fact that one needs more than cost-benefit analyses to get a decent grasp of actual human behavior. It was both a pleasure and an honor to share this insight with our students on many occasions.

Of course, I would like to express my gratitude to all those who helped with the realization of my projects in one way or another. There are certainly too many to name them all, but some stand out prominently in this regard. Special thanks go to Veronika Harder, Nathalie Stelzer, Laura Bickel, Jan Deller, Ben Mouelhi, Menusch Khadjavi and Dominik Boddin.

Last but not least, thanks to my family for everything.

Contents

Preface

This cumulative dissertation, titled „Essays on Behavioral Economics: Empirical Studies on Risk, Morality and Framing", comprises three self-contained empirical papers. Two of which are single-authored with the third being joint work with my supervisor Prof. Dr. Dr. Ulrich Schmidt. For the latter, we split the work equally among us.

These papers are included in the dissertation as follows:

Chapter 2:
Jelschen, H. & Schmidt, U. (2019). Windfall Gains and House Money: The Effects of Endowment History and Prior Outcomes on Risky Decision-Making

Chapter 3:
Jelschen, H. (2020). Moral Judgement in Probabilistic Dilemmas: A Descriptive Analysis

Chapter 4:
Jelschen, H. (2020). Framing and Gender Effects in the Sender-Receiver Game

List of Tables

2 Windfall Gains and House Money:

The Effects of Endowment History and Prior Outcomes on Risky Decision–Making

Appendix B: Additional statistics

3 Moral Judgement in Probabilistic Dilemmas – A Descriptive Analysis

4 Framing and Gender Effects in the Sender-Receiver Game

List of Figures

1 Introduction

Behavioral economics is a multi-faceted field and essentially subsumes all academic approaches that try to identify, model and explain economic behavior being at variance with classical economic assumptions which traditionally rest on rationality postulates. To put it a little differently and polemically, behavioral economics treats economic agents as real people rather than idealized parts of a model with unrealistically demanding characteristics.

The studies included in this dissertation focus on three key aspects of human decision-making that foster behavior being at odds with the famous (or infamous) notion of a "homo oeconomicus", that, by assumption, "can think like Albert Einstein, store as much memory as IBM's Big Blue, and exercise the willpower of Mahatma Gandhi," as Thaler and Sunstein unmaskingly put it (2009, p. 7).

The first issue being addressed in chapter two is the question whether risk tolerance is a universal and stable property of individual preferences or not and what determines possible changes of risk attitude.

In a seminal contribution, Thaler and Johnson (1990) detected the existence of a house money effect, which is defined as an increase in risk tolerance after previous gains resulting from a risky activity. Subsequent studies used the term house money effect also in case of windfall gains, i.e., easily acquired money like show-up fees or initial endowments in experiments which does not result from a risky investment. The present study is to the best of our knowledge the first that disentangles the house money effect and windfall gains. We find a clear and systematic pattern that windfall gains increase risk tolerance. In contrast, the house money effect is far less ubiquitous and seems to require skewed lotteries and/or a large number of rounds played. We, therefore, conclude that a careful distinction between windfall gains and the house money effect is warranted in future research.

The third chapter also addresses the question how individuals behave in a risky environment and combines this with fundamental issues of matters of live and death that may arise when people inevitably have to choose between two or more adverse outcomes. In this regard, it adds another dimension to the decision, which constitutes the second of the aforementioned factors and I denote this as "morality". The third aspect that is also explicitly seized on in this chapter is the question whether different representations of the same facts can yield distinct responses of individuals. This notion of a framing effect has become highly prominent in behavioral economics research since its introduction by Kahneman and Tversky and their famous "Asian disease" example (1981).

Namely, this research employs a multitude of probabilistic versions of two iconic variants of the trolley dilemma. In four studies, subjects rated moral permissibility of action in the bystander case – where one can divert a train, letting one person die instead of five – and the footbridge case – where

one can sacrifice one person to save the five – with outcome probabilities being changed systematically. Results show that decreasing attractiveness of intervention yields a decreasing perceived moral permissibility of the intervention. Furthermore, a constant ratio of expected outcomes leaves moral permissibility ratings unchanged on aggregate if outcome probabilities are identical, whereas they display the emergence of a common ratio effect in the bystander, but not in the footbridge situation in case of asymmetric probabilities if these are manipulated intrapersonally. Additionally, probability framing does not seem to be of major importance and previous findings that moral permissibility of intervention is denoted higher in the bystander case are confirmed.

The fourth chapter is a study that combines issues of morality and framing. The moral decision that the subjects had to make in this experimental setup is whether to lie to their counterpart in order to increase their own material payoff while decreasing the other's, thus it addresses morality in a less drastic way than the research depicted in chapter three. Framing effects are investigated by additional payoffs resulting from different activities being depicted as either gains or losses.

More precisely, this research employs four variants of the standard sender-receiver game by Gneezy (2005), with outcome valence being varied systematically. Depending on the frame of the game, a deceptive message, if acted upon, resulted in a higher gain for the sender and a lower gain for the receiver, a lower loss for the sender and a lower gain for the receiver, a higher gain for the sender and a higher loss for the receiver or a lower loss for the sender and a higher loss for the receiver. Results show that framing has no effect on senders' decisions to lie on aggregate. Analyses with respect to gender point towards female and male subjects being influenced differently by the framing manipulation. Women show a higher propensity to lie to avoid a higher loss and behave less deceptively if doing so increases the receiver's loss, with this pattern being reversed for men.

Chapter five provides a short conclusion.

References

Gneezy, U. (2005). Deception: The Role of Consequences. *The American Economic Review*, 95(1), 384-394.

Thaler, R. H. & Johnson, E. J. (1990). Gambling with the House Money and Trying to Break Even: The Effects of Prior Outcomes on Risky Choice. *Management Science*, 36(6), 643-660.

Thaler, R. H. & Sunstein, C. R. (2009). Nudge: improving decisions about health, wealth, and happiness. Rev. and expanded ed. New York: Penguin Books.

Tversky, A. & Kahneman, D. (1981). The Framing of Decisions and the Psychology of Choice. *Science*, 211(4481), 453-458.

2 Windfall Gains and House Money: The Effects of Endowment History and Prior Outcomes on Risky Decision–Making

Abstract

In a seminal contribution, Thaler and Johnson (1990) detected the existence of a house money effect which is defined as an increase in risk tolerance after previous gains resulting from a risky activity. Subsequent studies used the term house money effect also in case of windfall gains, i.e., easily acquired money like show-up fees or initial endowments in experiments which does not result from a risky investment. The present study is to the best of our knowledge the first that disentangles the house money effect and windfall gains. We find a clear and systematic pattern that windfall gains increase risk tolerance. In contrast, the house money effect is far less ubiquitous and seems to require skewed lotteries and/or a large number of rounds played. We, therefore, conclude that a careful distinction between windfall gains and the house money effect is warranted in future research.

2.1 Introduction

Despite classic normative theory assuming rational agents to base their decisions solely on an evaluation of incremental costs and benefits, it has become a widely supported insight that precedent events, outcomes and decisions may have the power to influence actual economic decision-making (e.g., Staw, 1976; Thaler, 1980; Arkes and Blumer, 1985).

Regarding decisions under risk and uncertainty a seminal contribution towards understanding behavior and changes of behavior in multi-round decisions involving monetary gains and losses was put forward by Thaler and Johnson (1990). Based on their experimental observations, they proposed Quasi-Hedonic Editing (QHE) as the underlying process in such contexts. On the one hand, according to QHE, individuals tend to integrate possible future losses with prior gains, i.e., regard them as reductions of previous gains until they are depleted rather than actual losses. In contrast, possible future gains are segregated from previous ones in this process. Assuming a Prospect Theory-like value function (concave in the gain domain, convex in the loss domain, and steeper for losses than for gains (Kahneman and Tversky, 1979)), this leads to enhanced risk proneness or at least mitigated risk aversion after a gain, which is called the house money effect. On the other hand, after a loss, individuals are assumed to segregate both possible future losses and gains from previous losses, which decreases risk proneness or amplifies risk aversive behavior. One vital exception are situations where substantial risk-taking offers the opportunity to break even, i.e., to offset prior losses.

These effects of prior outcomes in risky environments have been studied extensively in trading behavior in financial markets (Coval and Shumway, 2005; Frino et al. 2008; Liu et al., 2010; Hsu and Chow, 2013; Huang and Chan, 2014; Wen et al., 2014), lottery and portfolio choices (Weber and Zuchel, 2005) as well as in game show behavior (Gertner, 1993). However, in experimental economics, the term house money has also been used synonymously for easily gotten money in general and not been limited to situations where an individual is actually “gambling while ahead” as originally framed by Thaler and Johnson. Notably, the obligatory initial payment of monetary endowments in experimental settings is often referred to as house money (e.g., Clark, 2002; Ackert et al., 2006; Bosch-Domenech and Silvestre, 2010; Davis et al., 2010; Dannenberg et al., 2012; Rosenboim and Shavit, 2012; Cardenas et al., 2014; Corgnet et al., 2015) and also possibly fosters risk-taking behavior due to a generally higher marginal propensity of consumption (Arkes et al., 1994).

This paper draws a clear distinction between extraordinary riskless gains, such as initial payments in experiments, denoted as windfall money, and any positive difference between the current stake and the initial stake that has been acquired by previously taking risk, the house money.[1] We argue that there is no apparent reason to assume a priori that individuals treat these types of money identically, especially whenever an increase of one’s assets included the possibility of monetary losses before.

[1] It is worth noting that Thaler and Johnson (1990) do not draw this distinction.

Additionally, not making this distinction unnecessarily exacerbates any endeavor to identify the actual determinants of risk attitude and risk attitude changes in our view. For example, not observing one's risk proneness to increase after being lucky in a risky environment does not necessarily exclude the existence of a house money effect if the money put at stake was windfall money in the first place and therefore already induced behavior representing the maximum level of individual risk tolerance.

To the best of our knowledge, we report the first experimental results of an approach to disentangle the effects of prior riskless and risky gains on risk-taking behavior and risk attitude changes. In a first step, we aim at varying the extent to which subjects regard the experimental endowment as their own money rather than windfall money. As we cannot let people risk their own money and possibly leave the experiment in debt, we employ two different mechanisms to create the sensation of putting their own money at stake. Our baseline treatment does not exhibit such manipulation and includes a payment of experimental endowment in connection with a subsequent two round gambling task involving monetary gains and losses to elicit risk attitude and risk attitude changes. Our first manipulation involves a temporal separation of paying the endowment and the actual gambling task to let the money become part of the subjects' own assets in the course of time. This approach builds on contributions by Gourville and Soman (1998) and Shafir and Thaler (2006), who argue that money in certain mental accounts[2] depreciates over time, and has been employed effectively in the context of decisions under risk by Bosch-Domenech and Silvestre (2010), Rosenboim and Shavit (2012) as well as by Cardenas et al. (2014).

In a second manipulation, participants earn their endowment for the gambling task by completing questionnaires, while being fully informed that they would be compensated for their effort with a fixed payment. Although the effect of earned vs. windfall money has been studied in various domains, including charitable giving (Reinstein and Riener, 2012; Carlsson et al., 2013), dictator games (Cherry et al., 2002; Cherry and Shogren, 2008), public goods games (Cherry et al., 2005; Kroll et al., 2007) and experimental asset markets (Corgnet et al., 2015)[3], evidence on how it influences individual decisions under risk is surprisingly scarce. Zeelenberg and van Dijk (1997) contribute to this question by presenting results from hypothetical choices that "behavioral sunk costs", i.e., effort that has been exerted to earn money, can decrease subsequent risk proneness.

In total, our contribution is threefold. First, we elicit how the endowment's history influences individual risk attitude. Second, we examine how outcomes of initial decisions under risk change risk attitude. Third, we check whether risk attitude changes after losses and gains are influenced by the endowment's history.

[2] For an overview on mental accounting, see Thaler (1999).

[3] The authors also separate in time earnings task and market experiment in the same treatment.

The remainder of this paper is organized as follows. Chapter 2 illustrates our design and procedures of the experiment. Results are presented in Chapter 3 while Chapter 4 discusses open questions and concludes.

2.2 Design and procedure

We used a two round lottery game to elicit individual risk attitude which is a modified version of Gneezy and Potter's design (1997, 2003) as employed by Weber and Zuchel (2005). Participants were students enrolled in introductory economics and intermediate microeconomics courses at Kiel University, Germany. In total, 241 subjects participated in four different treatments. All treatments were run as pen and paper tasks in a classroom. The gambling task was designed as follows. All students were endowed with 8€ and assigned an identification number between 1 and n in the beginning. This number was also used for determining outcomes later on. In both of the two rounds, participants were allowed to buy lottery tickets, with the maximum number being limited to 10 units per round. One ticket cost 0.4€ and won or lost with equal probability. Each ticket paid 1€ in case of winning and nothing otherwise. Outcomes were perfectly positively correlated within subjects, i.e., all tickets bought by one person either won or lost. So participants effectively had to choose twice between the lotteries presented in Table 1 which are represented as gains and losses relative to the initial endowment in round 1 and the current stake in round 2 respectively.

Table 1: Menu of implicit lottery choices

No. of tickets	Resulting lottery	No. of tickets	Resulting lottery
0	0		
1	(0.6€, 0.5; -0.4€, 0.5)	6	(3.6€, 0.5; -2.4€, 0.5)
2	(1.2€, 0.5; -0.8€, 0.5)	7	(4.2€, 0.5; -2.8€, 0.5)
3	(1.8€, 0.5; -1.2€, 0.5)	8	(4.8€, 0.5; -3.2€, 0.5)
4	(2.4€, 0.5; -1.6€, 0.5)	9	(5.4€, 0.5; -3.6€, 0.5)
5	(3.0€, 0.5; -2.0€, 0.5)	10	(6.0€, 0.5; -4.0€, 0.5)

As our goal was to observe changes in risk attitude following a gain or a loss, the outcome of the first round was known before subjects made their decisions for the second round. The number of tickets bought serves as our measurement of individuals' risk attitude with a larger number representing a lower degree of risk aversion or even risk neutrality or risk proneness[4]. A coin toss determined whose tickets won after each of the two ticket buying decisions depending on whether the participant's assigned identification number was even or odd. By this mechanism, we ensured the groups of first-

[4] In fact, due to the lotteries' positive expected values, this method does not allow to distinguish between risk neutral and risk loving behavior as in both cases 10 units would be bought in each round.

round winners and losers to be of similar size. The last part of the experiment was a short questionnaire, including a question about the reasons why subjects made their ticket buying decisions. We are confident that subjects fully understood the incentivization mechanism of the lotteries and have no reason to assume otherwise.

All subjects were paid the money they won in addition to their initial endowment (a maximum of 12€) or had to pay back their losses (a maximum of 8€) at the very end of the experiment, i.e., individuals were confronted with paper gains/losses after round 1 that would not be realized before the end of round 2. The gambling task took around 15-20 minutes in most sessions.

Baseline treatment

For the baseline treatment (n=60, 33 male, 27 female), students were approached right after the tutorials that accompany the respective lectures. They were asked to participate in an experiment dealing with economic decision making. All subjects were then handed an envelope with 8€ and requested to check whether the money was actually in it. We did so to ensure that all subjects experienced the same sensation of actually holding the money in hand. They were subsequently informed about the procedure of the two round lottery game, both verbally and in writing. It was explicitly stated that no such thing as a 'correct' behavior existed. Both rounds of the gambling task were played subsequently and paid out at the end of the session.

Time treatment 1

To study the effect of time on risk-taking, in two treatments we separated the payment of the initial endowment from actual play of the two round lottery. In the first of these time treatments (n=64, 33 male, 31 female), the procedure was the same as for the baseline treatment, except that students were endowed with 8€ one week before they could bet this money in the lotteries. No decision had to be made at the time of the initial payment and participants were not instructed about the experiment's design at that point but merely informed that they could gamble with the 8€ in the tutorial one week later.

Time treatment 2

The second time treatment (n=59, 33 male, 26 female) was designed to yield an intrapersonal measurement of risk attitude changes over time. In contrast to Time 1, participants were informed about the task at the moment they received their initial endowment and were asked to make their ticket buying decision for the first round, though it would be played one week later. During the second meeting, they had to decide again on the number of tickets to buy in the first round, with this second decision being binding. They were not informed about this chance for revision in the first meeting. The second round of the gambling task as well as the payments were performed directly afterwards.

In both time treatments, participants were allowed to pay their losses in one of the following weeks in case they did not have enough money at hand and they were explicitly informed of this.

Work treatment

To capture the effect of behavioral sunk costs, i.e., effort to earn the endowment, on risk-taking, students in the work treatment (n=58, 28 male, 30 female) were recruited to complete two questionnaires and they were told to be compensated with 8€ for doing so. It took around 30 minutes for most of them. After receiving their payment, they were introduced to the two round lottery game that was played subsequently just as in the baseline treatment.

An overview of the timing and order of the events in all treatments along with the experimental instructions can be found in Appendix A.

2.3 Results

We discarded from the analysis all observations in the time treatments from subjects who did not show up for the second meeting (6 out of 70 in *Time 1,* 10 out of 69 in *Time 2*)[5].

2.3.1 Treatment effects on first round risk-taking behavior

To analyze the effects of the different treatments on risk attitude, we simply compare the ticket buying decisions in the first round across treatments, including the first (non-binding) decision made by subjects in *Time 2* (referred to as round 0 henceforth). Table 2 depicts the means of ticket buying decisions in round 0 and 1 for all four treatments (standard deviations in parentheses).

Table 2: Mean number of tickets bought by treatments

Treatment	Observations	Mean Round 0	Mean Round 1
Baseline	n=60	/	6.25 (2.61)
Time 1	n=64	/	5.31 (3.12)
Time 2	n=59	5.69 (2.73)	6.00 (2.85)
Work	n=58	/	4.10 (2.74)

We employ non-parametric Mann-Whitney U tests[6] to analyze possible treatment effects on risk-taking behavior. Table 3 depicts the respective significance levels for the differences between all reasonable pairings of ticket buying decisions. We exclude comparisons of second round decisions across treatments due to an overall lack of interpretability of those results brought about by different shares of winners and losers and confounding income effects resulting from different decisions in round 1.

[5] Drop out rates in the time treatments do not differ significantly (p=0.2741).

[6] Employing t tests instead does not change the qualitative results substantially. The respective significance levels are depicted in Table 8 in Appendix B.

Table 3: Differences in risk attitude between treatments

	Baseline	Time 1	Time 2/0	Time 2/1
Time 1	$p = 0.0285$**	-		
Time 2/0	$p = 0.2796$	$p = 0.3070$	-	
Time 2/1	$p = 0.6976$	$p = 0.1019$	$p = 0.3412$	-
Work	$p = 0.0000$***	$p = 0.0780$*	$p = 0.0031$***	$p = 0.0003$***

*Note: Mann-Whitney U tests; *=p<0.1, **=p<0.05, ***=p<0.01*

The results suggest a clear ranking between the treatments Baseline, Time 1 and Work. The degree of risk aversion appears to be lowest in the baseline treatment, with differences being statistically significant for a comparison with Time 1 (p<0.05) and even more so for a comparison with Work (p<0.01). Additionally, the number of tickets bought in Time 1 is significantly larger than in Work (p<0.1), although only at the 10% level. Round 0 and round 1 decisions in Time 2 were not significantly different from the baseline treatment or Time 1 at any conventional level, with the means being located between those of the other two treatments. Differences between these decisions in the second time treatment and Work are both found highly significant (p<0.01). It is worth noting that the revised decision of round 1 in Time 2 almost differs significantly from the decision of round 1 in Time 1 (p=0.1019), although the only difference between the treatment was an additional non-binding prior decision in Time 2 that could be changed at no (monetary) cost when making the actual ticket buying decision for round 1. Finally, initial and revised decisions in round 1 of Time 2 also do not differ significantly (p=0.3412), while the latter point at a lower degree of risk aversion as slightly more tickets are bought.

We also ran a Tobit regression to check for possible effects of gender, age and whether or not subjects engage at least occasionally in gambling outside the experiment. Using first round decisions, we pooled observations across all four treatments for this analysis. Results show that males take on significantly more risk than females (p<0.01). Additionally, they suggest that the number of tickets bought depends positively on subjects' age (p<0.05) as well as with taking part in gambling outside the experiment (p<0.05).

2.3.2 Effects of prior outcomes on risk-taking behavior

To analyze how winning in the first round affects risk attitude in the second round, we compare ticket buying decisions in the first round of those who win with their second decision. Similarly, we check whether a loss in the first round affects risk attitude in the second round by comparing decisions of first round losers in both rounds. We exclude those subjects from the analysis who did not buy a positive amount of tickets in the first round (none in Baseline, five in Time 1, three in Time 2 and nine in Work) and therefore have not experienced an actual gain or loss before making their second round

decision. Additionally, we discarded all subjects from the sample who bought the maximum amount of ten tickets in the first round (17 in Baseline, 15 in Time 1, 14 in Time 2 and 5 in Work) to counteract a possible downward bias of risk attitude change assessment.[7]

Table 4 depicts the means of tickets bought in all rounds for winners and losers in all treatments (standard deviations in parentheses) as well as the significance levels that decisions in round 1 differ from decisions in round 2 according to non-parametric Wilcoxon tests.[8] In the first two rows of the table, we also include a comparison of pooled decisions by first round winners and losers from treatments Baseline, Time 1 and Work, while excluding observations in Time 2 from this part of the analysis. As it will become obvious in the remainder of this section, Time 2 decisions stand out in several ways and cannot be reasonably included in a combined evaluation.

Table 4: Differences in risk attitude between rounds within treatments

Treatment	Observations	Mean / Round 1	Mean / Round 2	p-value
All (except Time 2)/winners	n=65	4.43 (1.37)	4.37 (2.93)	0.3603
All (except Time 2)/losers	n=66	4.47 (1.56)	4.70 (3.18)	0.8682
Baseline /winners	n=22	4.73 (1.28)	5.18 (2.58)	0.8102
Baseline /losers	n=21	4.81 (1.33)	5.57 (3.08)	0.3228
Time 1/winners	n=22	4.41 (1.40)	3.86 (3.11)	0.3023
Time 1/losers	n=22	4.23 (1.66)	4.45 (2.96)	0.8908
Time 2/winners	n=19	4.68 (1.49)	2.84 (1.57)	0.0008***
Time 2/losers	n=23	5.35 (1.61)	4.26 (2.47)	0.0180**
Work/winners	n=21	4.14 (1.42)	4.05 (3.04)	0.3921
Work/losers	n=23	4.39 (1.64)	4.13 (3.43)	0.5290

Note: Wilcoxon tests; *=p<0.1, **=p<0.05, ***=p<0.01

Figure 1a gives a graphical representation of decisions for first round winners while Figure 1b depicts the respective results for first round losers. For treatments *Baseline*, *Time 1* and *Work,* the analysis suggests no significant effect of the first round outcome on risk-taking behavior in the second round of the lottery. This holds true both for an evaluation of those who experienced a gain as well as for those who lost in the first round. Treatment *Time 2* is different since here we find a significant increase of risk aversion in round 2 for both first round winners ($p<0.01$) and first round losers ($p<0.05$). This result will be discussed below.

[7] Retaining all maximum buyers in the analysis does not considerably change the following qualitative results.
[8] Employing t tests instead does not change the qualitative results substantially. The respective significance levels are depicted in Table 9 in Appendix B.

Figure 1a: Mean numbers of tickets bought / first round winners

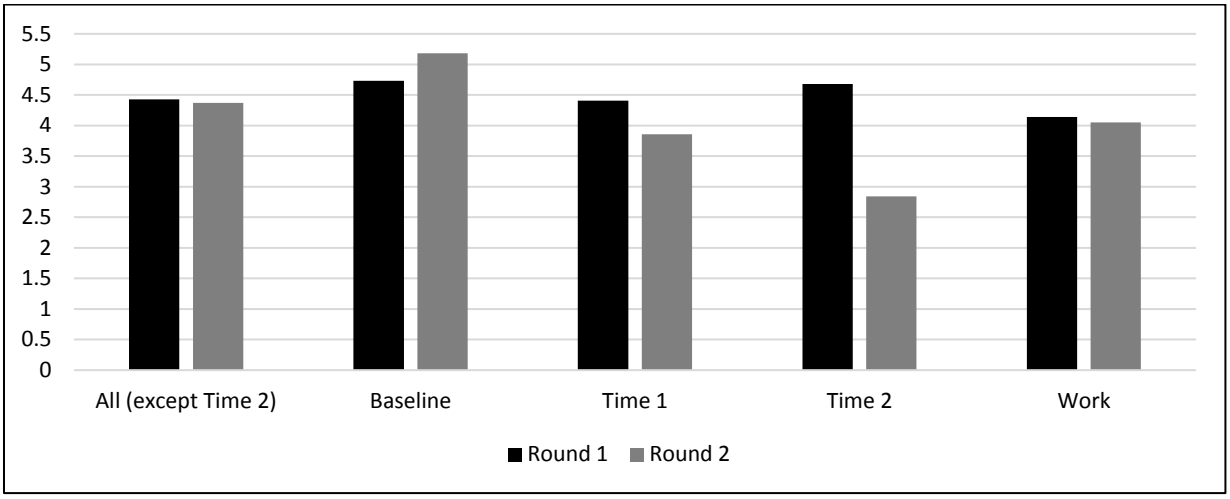
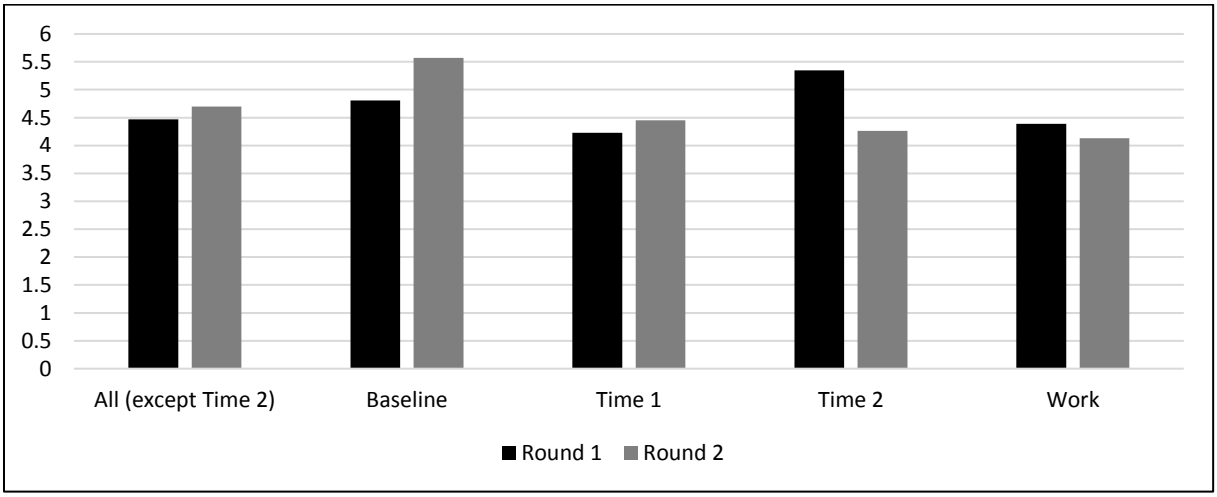


Figure 1b: Mean numbers of tickets bought / first round losers



Having established that neither first round winners nor first round losers seem to be affected by first rounds' outcomes in a significant and predictable way (except in Time 2), we take a closer look at possible differences in reactions to winning or losing within treatments. We therefore compare cross –subject behavioral changes (the changes in numbers of tickets bought) of first round winners with those of first round losers and find that these differences too are not significant at any conventional level, now also in Time 2. Significance levels of two-sample, two-sided t tests[9] are given as p=0.73 in Baseline, p=0.32 in Time 1, p=0.20 in Time 2 and p=0.87 in Work.

We now turn to an investigation of how the magnitude of gains or losses in the first round affects behavior in round 2. We use OLS regressions to quantify the impact of relative earnings, i.e., any

[9] The results change only slightly by employing non-parametric Wilcoxon tests instead. P values are then given by 1.000 (*Baseline*), 0.2197 (*Time 1*), 0.0968 (*Time 2*) and 0.9440 (*Work*).

deviation from the 8€ endowment after round 1, on the change of the amount of tickets bought between rounds. Table 5 shows coefficient values and significance levels of these regressions for winners and losers combined in each treatment as well as for pooled observations of all treatments except Time 2. Across columns, Table 5 also depicts main results of separate regressions conditional on winning or losing for all treatments as well as for pooled observations (again, excluding subjects in Time 2).

Table 5: Effect of relative earnings on differences in numbers of tickets bought

Treatment	Winners	Losers	Combined
All (except Time 2)	$\beta = -0.67$ $p = 0.125$	$\beta = 0.70$ $p = 0.244$	$\beta = -0.075$ $p = 0.498$
Baseline	$\beta = -0.93$ $p = 0.216$	$\beta = 1.91$ $p = 0.164$	$\beta = -0.06$ $p = 0.729$
Time 1	$\beta = -0.41$ $p = 0.606$	$\beta = -0.73$ $p = 0.300$	$\beta = -0.21$ $p = 0.204$
Time 2	$\beta = -1.17^{**}$ $p = 0.011$	$\beta = 0.01$ $p = 0.984$	$\beta = -0.20^{*}$ $p = 0.075$
Work	$\beta = -0.93$ $p = 0.257$	$\beta = 1.62$ $p = 0.162$	$\beta = 0.04$ $p = 0.848$

*Note: OLS regressions; *=p<0.1, **=p<0.05, ***=p<0.01*

As in the binary win/loss analysis, we do not find significant effects of relative earnings at any conventional level except for winners and the combined evaluation in *Time 2*. By looking at the first row, we can at best detect an overall tendency that subjects react more cautiously as previous gains or losses increase, as captured by a decreasing change in numbers of tickets bought in both cases. However, additional to emphasizing the lack of significance, some further words of extreme caution are in order when interpreting these results. As the number of tickets that can be bought is limited, high first round gains or losses entail a reduction of options to increase the number of tickets in the second round compared to a low previous gain or loss, although they leave the choice set unchanged (a maximum of 10 tickets can be bought in round two, regardless of the first round outcome). For example, anyone making her second round decision after a gain of 3€ (implying 5 tickets were bought in the first round) can increase or decrease the number of tickets by 5 units each, while after a 1.2€ gain one can increase it by 8 units but decrease it by only 2. In fact, assuming that subjects just randomly pick one of their options in round 2 independently of the previous outcome, the same pattern as reported above would emerge.

Finally, we investigate if winning or losing in the first round effects second round behavior differently in the four treatments. We therefore employ pairwise comparisons of behavioral changes

(the changes in numbers of tickets bought) for both, first round winners and first round losers between treatments. Table 6 shows significance levels of two-sample, two-sided t tests[10].

Table 6: Behavioral changes between treatments conditional on first outcome

	Baseline	Time 1	Time 2
Time 1	winners: $p = 0.240$ losers: $p = 0.525$	-	
Time 2	winners***: $p = 0.002$ losers**: $p = 0.029$	winners*: $p = 0.095$ losers**: $p = 0.032$	-
Work	winners: $p = 0.533$ losers: $p = 0.322$	winners: $p = 0.627$ losers: $p = 0.575$	winners**: $p = 0.035$ losers: $p = 0.331$

*Note: Two-sided t-tests; *=p<0.1, **=p<0.05, ***=p<0.01*

We do not observe any significant differences for comparisons between winners and losers in *Baseline*, Time 1 and *Work,* suggesting that their reaction does not depend on the initial endowment's history in a predictable way. The only significant differences occur if we compare reactions of winners and losers in *Time 2* with those in the remaining three treatments, which does not come as a surprise in the light of the previously reported results. Table 7 depicts the share of subjects who changed their behavior between rounds.

Table 7: Share of different decisions between rounds

Treatment		**0→1**			**1→2**		
		>	**<**	**/**	**>**	**<**	**/**
All	combined (n=173)				23.7%	41.6%	34.7%
	winners (n=84)				20.2%	44%	35.7%
	losers (n=89)				27%	39.3%	33.7%
Baseline	combined (n=43)				34.9%	27.9%	37.2%
	winners (n=22)				27.3%	27.3%	45.5%
	losers (n=21)				42.9%	28.6%	28.6%
Time 1	combined (n=44)				27.3%	40.9%	31.8%
	winners (n=22)				31.8%	50%	18.2%
	losers (n=22)				22.7%	31.8%	45.5%
Time 2	combined (n=42)	23.8%	14.3%	61.9%	7.1%	54.8%	38.1%
	winners (n=19)				0%	63.2%	36.8%
	losers (n=23)				13%	47.8%	39.1%
Work	combined (n=44)				25%	43.2%	31.8%
	winners (n=21)				19%	38.1%	42.9%
	losers (n=23)				30.4%	47.8%	21.7%

[10] Employing non-parametric Wilcoxon tests instead does not change the qualitative results substantially. The respective significance levels are depicted in Table 10 in Appendix B.

These descriptive results provide further evidence for a lack of a clear pattern in how first round outcomes influence subsequent decisions and can be backed up by additional results from simple Probit regressions. We check whether the likelihood of observing a winner is significantly depending on observing that the number of tickets bought has increased in round 2 and find that it does not at any conventional significance level (p=0.284 for *Baseline,* p=0.499 for *Time 1,* p=0.383 for *Work* and p=0.298 for pooled observations). *Time 2* constitutes a special case here, as none of the first-round winners actually increase the number of tickets bought. Around one third of subjects in all treatments do not change their behavior after round 1 at all. In *Baseline*, *Time 1* and *Work*, we observe a considerable number of subjects changing behavior in both directions, which holds for both first round winners and losers with the direction of the changes being ambiguous. In *Time 2*, none of the winners and only three losers buy more tickets in the second round. Notably, only around 40% changed their initial decision (round 0) in *Time 2*. So, the insignificant effects of first round outcomes on average risk-taking behavior in the former three treatments might be driven by these effects being ambiguous rather than nonexistent at all.

To gain further insight in this regard and to test whether the behavioral changes are just due to some kind of general noise (i.e., not causally linked to a previous outcome at all), we ran an additional treatment (n=33, 16 male, 17 female), that closely resembled the protocol of the *Baseline* treatment. The only difference was that subjects would not learn about the first round's outcome before they would have to make their decision for round two, which was of course common knowledge.

We find that decisions do not differ significantly between rounds at any conventional level, confirmed by both Wilcoxon and t tests. Again excluding all non-buyers and all maximum-buyers from the analysis as before, we find that 50% do not change their decision between rounds, while 25% increase and 25% decrease their number of tickets bought. This fraction of non-changers is the highest we observe overall, yet not significantly different from those in the other treatments at any conventional level. However, if we check whether the intensity of reactions after round one increases if subjects learn about the first round outcome between rounds, i.e., if we compare the average absolute behavioral changes defined as the absolute values of the differences of number of tickets bought between rounds, we find that this is lower in this additional treatment compared to *Baseline, Time 1 and Work* according to one-sided t tests (*Baseline*: p=0.0304, *Time 1*: p=0.0568, *Work*: p=0.0132).

With all due care, we conclude here that much of the variation between rounds can be reasonably attributed to noise, but it seems legit to infer that winning or losing in round one amplifies behavioral changes between rounds on average.

2.4 Discussion and conclusion:

On the one hand, our results suggest the effectiveness of our treatment manipulation with respect to risk attitude. Giving away the endowment as windfall money directly before the decision obviously induces the strongest willingness to gamble, while earning the endowment makes individuals most reluctant to do so. The passage of time seems to dampen the risk proneness enhancing effect of being granted windfall money, with this effect not being strong enough to outright resemble the situation in the work treatment. We interpret this as unambiguous evidence for the existence of a windfall money effect with respect to risk attitude. Additional support for this interpretation is found in participants' post-experiment answers to the question of the considerations which guided their decisions. In the baseline treatment, ten subjects actually stated explicitly that they had the feeling that they were not actually gambling with their own money, while five in time treatment 1, seven in time treatment 2 and only one in the work treatment made such statements.

On the other hand, we fail to identify any systematic pattern of risk attitude changes following a first round gain or loss in the baseline treatment, time treatment 1 and the work treatment, challenging the existence of a house money effect as defined in our analysis, regardless of the initial endowment's history.

Observations in time treatment 2 constitute a remarkable exception in this regard, as individuals show substantially increased risk averse behavior in round 2, both after winning and losing in round 1. Several deliberations can be employed to shed some light on these peculiar results, although all of them should be treated with caution due to their speculative nature. Recall that subjects' decisions in round 0 do not significantly differ from those in the baseline treatment for round 1.[11] This should not come as a surprise, as individuals cannot reasonably be expected to actually include a correct forecast of their risk attitude change as observed in time treatment 1. It seems permissible to assume that this first decision serves as an anchor (Tversky and Kahneman, 1974) for the second decision when the actual gambling is performed (56 out of 59 subjects stated afterwards that they remembered their decision in round 0 when making their decision in round 1).

While there is no plausible reason to assume that individuals' "true" risk attitudes differ significantly from the ones in time treatment 1, a majority of subjects do not revise their decision and buy more tickets in round 1 than they presumably would have without this anchor. This may be the result of a generally effective status quo bias (Samuelson and Zeckhauser, 1988), i.e., a general inertia or reluctance to change one's previously made decision. This inertia may be amplified by an experimenter demand effect (Zizzo, 2010), as modalities of the gambling task have not changed and

[11] Although not within the scope of this paper, note that this implies that we do not observe evidence for systematically distinct discount rates for possible future gains and losses in this decision (see Ahlbrecht and Weber, 1997).

subjects feel obliged to stick to their initial decision, in order not to appear "inconsistent" or "irrational" in their own view. Six participants rationalized their sticking to their first decision by explicitly stating that there "had been no reason to change" in the post-experiment questionnaire. In fact, subjects may be even inadvertently pushed into not changing their decision as they were explicitly told that "the conditions had not changed from last week's". Nevertheless, once the first round is played, subjects are relieved of that aforementioned anchor when making their decision in round 2 and free to choose according to their "true" risk attitude. This effect may be enhanced by people feeling the necessity to make up for the "overgambling" in round 1 by acting more cautiously in the following, resulting in a significant decrease of number of tickets bought. In sum, our manipulation in the second time treatment did not enable us to assess an intrapersonal measure of reduced risk proneness as originally intended but yielded some insights that are noteworthy in their own right.

Recently, attention has been drawn to possibly different effects of paper gains/losses vs. realized gains/losses. Davis et al. (2010) report that paying a show-up fee after the experiment rather than before decreases risky behavior, measured as refraining from information purchases in their experiment. Reinstein and Riener (2012) show that donations in a dictator game decrease substantially if the endowment to be allocated is handed to recipients in cash instead of being shown on a computer screen. Imas (2016) observes individuals to take on greater risk after a paper loss while taking on less risk after a realized loss.

Our design includes both the tangibility of endowment paid at the beginning and paper losses/gains after round 1, as the final payments were not realized before the end of round 2. It is worth noting at this point that we neither observe an increase in risk-taking behavior after paper losses, nor after paper gains that had been generated by betting money that participants actually held at hand. Our results on paper losses are in contrast to those of Imas (2016) who observes a significant increase in risk-taking after paper losses with a rather similar experimental design. The only difference to our design is that Imas uses skewed lotteries and four instead of two investment rounds. Also Langer and Weber (2008) observe increased risk-taking after paper losses in a design with skewed lotteries and multiple (namely 30) rounds. Since we have about twice the sample size of Study 1 of Imas our failure to detect a house money effect cannot be attributed to an underpowered study but should be due to the design differences. The question whether and how the skew of lotteries, the number of rounds or the combination of both can lead to the emergence of a house money effect is left for future research. Systematically varying prospects' payoffs and probabilities as well as the number of rounds may shed more light on necessary presuppositions for risk-taking to change systematically after gains or losses. Therefore, it should be emphasized here that we do not claim that our results rule out the existence of such an enhancing effect on risk-taking propensity of prior risky gains per se, but rather

that this phenomenon is certainly far from being reasonably considered a behavioral regularity, as we do not observe such systematic behavior in any of our treatments.

While we cannot detect a predictable change of risk-taking behavior by first round losers, caution is in order when interpreting this result. Our findings are not generally at variance with the concept of a break-even effect, i.e., an inclination to take on greater risk to equalize prior losses if possible. In fact, "breaking even" is always possible in our experiment without changing behavior between rounds and even by reducing the number of tickets bought whenever a minimum of three tickets was bought in round 1.

A worthwhile question to address in future research is how long the time span between paying the endowment and the actual task should be and if a systematic variation of this length would yield systematically different results. For the moment, time spans in the literature are extremely different (three days in Corgnet et al., 2015; one week in our study; two weeks in Rosenboim and Shavit, 2012; three weeks in Cardenas et al., 2014; and four months in Bosch-Domenech and Silvestre, 2010) and seem determined arbitrarily, but all still yield significant effects.

Concluding, our results suggest carefully differentiating between a windfall gain effect and the house money effect. They should also intensify the awareness that paying experimental endowments as windfall money may significantly decrease observable risk aversion.

References

Ackert, L. F., Charupat, N., Church, B. K. & Deaves, R. (2006). An experimental examination of the house money effect in a multi-period setting. *Experimental Economics*, 9, 5-16.

Ahlbrecht, M. & Weber, M. (1997). An Empirical Study on Intertemporal Decision Making under Risk. *Management Science*, 43(6), 813-826.

Arkes, H. R. & Blumer, C. (1985). The Psychology of Sunk Cost. *Organizational Behavior and Human Decision Processes*, 35, 124-140.

Arkes, H. R., Joyner, C. A., Pezzo, M. V., Nash, J. G., Siegel-Jacobs, K. & Stone, E. (1994). The Psychology of Windfall Gains. *Organizational Behavior and Human Decision Processes*, 59, 331-347.

Bosch-Domenech, A. & Silvestre, J. (2010). Averting risk in the face of large losses: Bernoulli vs. Tversky and Kahneman. *Economic Letters*, 107, 180-182.

Cardenas, J. C., De Roux, N., Jaramillo, C. R. & Martinez, L. R. (2014). Is it my money or not? An experiment on risk aversion and the house-money effect. *Experimental Economics*, 17, 47-60.

Carlsson, F., He, H. & Martinsson, P. (2013). Easy come, easy go. The role of windfall money in lab and field experiments. *Experimental Economics*, 16, 190-207.

Cherry, T. L. & Shogren, J. F. (2008). Self-interest, sympathy and the origin of endowments. *Economics Letters*, 101, 69-72.

Cherry, T. L., Frykblom, P. & Shogren, J. F. (2002). Hardnose the Dictator. *American Economic Review*, 92(4), 1218-1221.

Cherry, T. L., Kroll, S. & Shogren, J. F. (2005). The impact of endowment heterogeneity and origin on public good contributions: evidence from the lab. *Journal of Economic Behavior and Organization*, 57, 357-365.

Clark, J. (2002). House Money Effects in Public Good Experiments. *Experimental Economics*, 5, 223-231.

Corgnet, B., Hernan-Gonzalez, R., Kujal, P. & Porter, D. (2015). The Effect of Earned Versus House Money on Price Bubble Formation in Experimental Asset Markets. *Review of Finance*, 19, 1455-1488.

Coval, J. D., and Shumway, T. (2005). Do Behavioral Biases Affect Prices? *The Journal of Finance*, 60(1), 1-34.

Dannenberg, A., Riechmann, T., Sturm, B. & Vogt, C. (2012). Inequality aversion and the house money effect. *Experimental Economics*, 15, 460-484.

Davis, L. R., Joyce, B. P. & Roelofs, M. R. (2010). My money or yours: house money payment effects. *Experimental Economics*, 13, 189-205.

Frino, A., Grant, J. & Johnstone, D. (2008). The house money effect and local traders on the Sydney Future Exchange. *Pacific-Basin Finance Journal*, 16, 8-25.

Gertner, R. (1993). Game Shows and Economic Behavior: Risk-Taking on "Card Sharks". *The Quarterly Journal of Economics*, 108(2), 507-521.

Gneezy, U. & Potters, J. (1997). An Experiment on Risk-taking and Evaluation Periods. *The Quarterly Journal of Economics*, May 1997, 631-645.

Gneezy, U., Kapteyn, A. & Potters, J. (2003). Evaluation Periods and Asset Prices in a Market Experiment. *The Journal of Finance*, 58(2), 821-837.

Gourville, J. T. & Soman, D. (1998). Payment Depreciation: The Behavioral Effects of Temporally Separating Payments from Consumption. *The Journal of Consumer Research*, 25(2), 160-174.

Huang, Y. C. & Chan, S. H. (2014). The house money and break-even effects for different types of traders: Evidence from Taiwan futures markets. *Pacific-Basin Finance Journal*, 26, 1-13.

Hsu, Y.-L. & Chow, E. H. (2013). The house money effect on investment risk-taking: Evidence from Taiwan. *Pacific-Basin Finance Journal*, 21, 1102-1115.

Imas, A. (2016). The Realization Effect: Risk-Taking after Realized versus Paper Losses. *American Economic Review*, 106(8), 2086-2109.

Kahneman, D. & Tversky, A. (1979). Prospect Theory: An Analysis of Decision under Risk. *Econometrica*, 47(2), 263-291.

Kroll, S., Cherry, T. L. & Shogren, J. F. (2007). The impact of endowment heterogeneity and origin on contributions in best-shot public good games. *Experimental Economics*, 10, 411-428.

Langer, T. & Weber, M. (2008). Does Commitment or Feedback Influence Myopic Loss Aversion? *Journal of Economic Behavior and Organization*, 67(3-4), 810–819.

Liu, Y.-J., Tsai, C.-L., Wang, M.-C. & Zhu, N. (2010). Prior Consequences and Subsequent Risk-taking: New Field Evidence from the Taiwan Future Exchange. *Management Science*, 56(4), 606-620.

Reinstein, D. & Riener, G. (2012). Decomposing desert and tangibility effects in a charitable giving experiment. *Experimental Economics*, 15, 229-240.

Rosenboim, M. & Shavit, T. (2012). Whose money is it anyway? Using prepaid incentives in experimental economics to create a natural environment. *Experimental Economics*, 15, 145-157.

Samuelson W. & Zeckhauser, R. (1988). Status Quo Bias in Decision Making. *Journal of Risk and Uncertainty*, 1, 7-59.

Shafir, E. & Thaler, R. H. (2006). Invest now, drink later, spend never: On the mental accounting of delayed consumption. *Journal of Economic Psychology*, 27, 694-712.

Staw, B. M. (1976). Knee-Deep in the Big Muddy: A Study of Escalating Commitment to a Chosen Course of Action. *Organizational Behavior and Human Performance*, 16, 27-44.

Thaler, R. H. (1980). Toward a Positive Theory of Consumer Choice. *Journal of Economic Behavior and Organization*, 1, 39-60.

Thaler, R. H. & Johnson, E. J. (1990). Gambling with the House Money and Trying to Break Even: The Effects of Prior Outcomes on Risky Choice. *Management Science*, 36(6), 643-660.

Thaler R. H. (1999). Mental Accounting Matters. *Journal of Behavioral Decision Making*, 12, 183-206.

Tversky, A. & Kahneman, D. (1974). Judgement under Uncertainty: Heuristics and Biases. *Science*, 185(4157), 1124-1131.

Wen, F., Gong, X., Chao, Y. & Chen, X. (2014). The Effects of Prior Outcomes on Risky Choice: Evidendence from the Stock Market. *Mathematical Problems in Engineering*, vol. 2014, Article ID 272518, 8 pages.

Weber, M. & Zuchel, H. (2005*).* How Do Prior Outcomes Affect Risk Attitude? Comparing Escalation of Commitment and the House-Money Effect. *Decision Analysis*, 2(1), 30-43

Zeelenberg, M. & van Dijk, E. (1997). A reverse sunk cost effect in risky decision making: Sometimes we have too much invested to gamble. *Journal of Economic Psychology*, 18, 677-691.

Zizzo, D. J. (2010). Experimenter demand effects in economic experiments. *Experimental Economics*, 13, 75-98.

Appendix A: Experimental instructions (translated from German)

Timing and order of the events in the treatments. The accompanying instructions in parantheses:

Baseline:

1) Payment of endowment
2) General instructions and decision for round one (*Baseline-1*)
3) Round one is played
4) Decision for round two (*Baseline-2*)
5) Round two is played.
6) Short questionnaire (*Baseline-3*)
7) Final payments.

Time 1:

First meeting:

1) Payment of endowment. Subjects were told that they would have the chance to gamble with this money in the following week.

Second meeting:

1) General instructions and decision for round one (*Time 1-1*)
2) Round one is played
3) Decision for round two (*Time 1-2*)
4) Round two is played
5) Short questionnaire (*Time 1-3*)
6) Final payments

Time 2:

First meeting:

1) Payment of endowment
2) General instructions and first decision for round one (denoted round zero in the text) (*Time 2-1*)

Second meeting:

1) Second decision for round one (*Time 2-2*)
2) Round one is played
3) Decision for round two (*Time 2-3*)
4) Round two is played
5) Short questionnaire (*Time 2-4*)
6) Final payments

Work:

1) Completing the questionnaires.
2) Payment for completing the questionnaires.
3) General instructions and decision for round one (*Work-1*)
4) Round one is played
5) Decision for round two (*Work-2*)
6) Round two is played
7) Short questionnaire (*Work-3*)
8) Final payments

Robustness Check:

1) Payment of endowment
2) General instructions and decision for round one (*Robustness Check-1*)
3) Round one is played (outcome is not announced)
4) Decision for round two (*Robustness Check-2*)
5) Outcome of round one is made known
6) Round two is played.
7) Short questionnaire (*Robustness Check-3*)
8) Final payments.

Instructions for *Baseline* treatment:

Baseline-1

Thank you for participating in this behavioral economics experiment.
Please answer the following questions about yourself:

I am male ____ female ____

I am _____ years old.

I study _____ in the _____. semester.

The experiment you are participating in consists of a two-stage gambling task.
In the beginning, you are endowed with an amount of 8€ that you can stake in two consecutive lotteries. You can keep the money that you have at the end of round two.

The characteristics of the lotteries are as follows:
- You can use your 8€ to buy lottery tickets in both of the two rounds.
- One lottery ticket costs 0.40€ and wins or loses with a probability of 50%.
- If your tickets win, you will receive 1€ per ticket.
- If your tickets lose, they are blanks and you do not receive any money for them.
- The outcomes of all tickets you buy are perfectly positively correlated, i.e., either all tickets win or all tickets lose.
- The number of tickets you can buy is limited to 10 per round.
- **Please note:** There is no such thing as a right or wrong behavior in this experiment. It is solely up to you, how many tickets you want to buy.

Please decide how many lottery tickets you want to buy in round one.

I buy _____ tickets in round one. (**Please note:** The number must be between 0 and 10)

Baseline-2

In the second round of the experiment, you have the possibility to buy lottery tickets again. The conditions are the same as in round one:

- One ticket costs 0.40€
- Your tickets win with a probability of 50% and you receive 1€ per ticket.
- Either all of your tickets win or all lose.
- The number of tickets you buy must be between 0 and 10.
- **Please note:** As in round one, you are completely free to choose how many tickets you want to buy.

Please decide how many tickets you want to buy in round two.

I buy _____ tickets in round two. (**Please note:** The number must be between 0 and 10)

Baseline-3

Please take some time to answer these questions:

1) Which considerations influenced your decision on how many tickets to buy (especially in round two)?

_____

_____

_____

_____

2) Do you regularly play games of chance (slot machines, lottery, sports bets, poker, etc.)? If so, how often?

yes___ no___

_____

_____

_____

3) Are you familiar with the term “House Money Effect“?

yes___ no___

Thank you for participation!

Instructions for *Time 1* treatment:

Time 1-1

Thank you for participating in this behavioral economics experiment.
Please answer the following questions about yourself:

I am male ____ female ____

I am _____ years old.

I study _____ in the _____. semester.

The experiment you are participating in consists of a two-stage gambling task.
Last week, you were endowed with an amount of 8€ that you can stake in two consecutive lotteries. You can keep the money that you have at the end of round two.

The characteristics of the lotteries are as follows:
- You can use your 8€ to buy lottery tickets in both of the two rounds.
- One lottery ticket costs 0.40€ and wins or loses with a probability of 50%.
- If your tickets win, you will receive 1€ per ticket.
- If your tickets lose, they are blanks and you do not receive any money for them.
- The outcomes of all tickets you buy are perfectly positively correlated, i.e., either all tickets win or all tickets lose.
- The number of tickets you can buy is limited to 10 per round.
- **Please note:** There is no such thing as a right or wrong behavior in this experiment. It is solely up to you, how many tickets you want to buy.

Please decide how many lottery tickets you want to buy in round one.

I buy _____ tickets in round one. (**Please note:** The number must be between 0 and 10)

Time 1-2

In the second round of the experiment, you have the possibility to buy lottery tickets again. The conditions are the same as in round one:

- One ticket costs 0.40€
- Your tickets win with a probability of 50% and you receive 1€ per ticket.
- Either all of your tickets win or all lose.
- The number of tickets you buy must be between 0 and 10.
- **Please note:** As in round one, you are completely free to choose how many tickets you want to buy.

Please decide how many tickets you want to buy in round two.

I buy _____ tickets in round two. (**Please note:** The number must be between 0 and 10)

Time 1-3

Please take some time to answer these questions:

1) Which considerations influenced your decision on how many tickets to buy (especially in round two)?

_____

_____

_____

_____

2) Do you regularly play games of chance (slot machines, lottery, sports bets, poker, etc.)? If so, how often?

yes___ no___

_____

_____

_____

3) Are you familiar with the term “House Money Effect“?

yes___ no___

Thank you for participation!

Instructions for *Time 2* treatment:

Time 2-1

Thank you for participating in this behavioral economics experiment.
Please answer the following questions about yourself:

I am male ____ female ____

I am _____ years old.

I study _____ in the _____. semester.

The experiment you are participating in consists of a two-stage gambling task.
In the beginning, you are endowed with an amount of 8€ that you can stake in two consecutive lotteries. These lotteries will be played next week. You can keep the money that you have at the end of round two.

The characteristics of the lotteries are as follows:
- You can use your 8€ to buy lottery tickets in both of the two rounds.
- One lottery ticket costs 0.40€ and wins or loses with a probability of 50%.
- If your tickets win, you will receive 1€ per ticket.
- If your tickets lose, they are blanks and you do not receive any money for them.
- The outcomes of all tickets you buy are perfectly positively correlated, i.e., either all tickets win or all tickets lose.
- The number of tickets you can buy is limited to 10 per round.
- **Please note:** There is no such thing as a right or wrong behavior in this experiment. It is solely up to you, how many tickets you want to buy.

Please decide how many lottery tickets you want to buy in round one.

I buy _____ tickets in round one. (**Please note:** The number must be between 0 and 10)

Time 2-2

Today, you have the possibility to decide again how many tickets you want to buy in round one. The conditions have not changed from last week‘s.

They are as follows:

- One ticket costs 0.40€
- Your tickets win with a probability of 50% and you receive 1€ per ticket.
- Your tickets lose with a probability of 50%.
- Either all of your tickets win or all lose.
- The number of tickets you buy must be between 0 and 10.
- **Please note:** As in round one, you are completely free to choose how many tickets you want to buy in round two.

Please decide how many tickets you want to buy in round one.

I buy _____ tickets in round two. (**Please note:** The number must be between 0 and 10)

Time 2-3

In the second round of the experiment, you have the possibility to buy lottery tickets again. The conditions are the same as in round one:

- One ticket costs 0.40€
- Your tickets win with a probability of 50% and you receive 1€ per ticket.
- Either all of your tickets win or all lose.
- The number of tickets you buy must be between 0 and 10.
- **Please note:** As in round one, you are completely free to choose how many tickets you want to buy.

Please decide how many tickets you want to buy in round two.

I buy _____ tickets in round two. (**Please note:** The number must be between 0 and 10)

Time 2-4

Please take some time to answer these questions:

1) Which considerations influenced your decision on how many tickets you buy (especially in round two)?

_____

_____

_____

_____

2) Do you regularly play games of chance (slot machines, lottery, sports bets, poker etc.) ? If so, how often?

yes___ no___

_____

_____

_____

3) Did you change your decision on how many tickets to buy in round one? Why?/Why not?

yes___ no___ don't know___

_____

_____

_____

4) Are you familiar with the term "House Money Effect"?

yes___ no___

Thank you for participation!

Instructions for *Work* treatment:

Work-1

Thank you for participating in this behavioral economics experiment.
Please answer the following questions about yourself:

I am male ____ female ____

I am _____ years old.

I study _____ in the _____. semester.

The experiment you are participating in consists of a two-stage gambling task.
You can stake the amount of 8€ that you have received for completing the questionnaires in two consecutive lotteries. You can keep the money that you have at the end of round two.

The characteristics of the lotteries are as follows:
- You can use your 8€ to buy lottery tickets in both of the two rounds.
- One lottery ticket costs 0.40€ and wins or loses with a probability of 50%.
- If your tickets win, you will receive 1€ per ticket.
- If your tickets lose, they are blanks and you do not receive any money for them.
- The outcomes of all tickets you buy are perfectly positively correlated, i.e., either all tickets win or all tickets lose.
- The number of tickets you can buy is limited to 10 per round.
- **Please note:** There is no such thing as a right or wrong behavior in this experiment. It is solely up to you, how many tickets you want to buy.

Please decide how many lottery tickets you want to buy in round one.

I buy _____ tickets in round one. (**Please note:** The number must be between 0 and 10)

Work-2

In the second round of the experiment, you have the possibility to buy lottery tickets again. The conditions are the same as in round one:

- One ticket costs 0.40€
- Your tickets win with a probability of 50% and you receive 1€ per ticket.
- Either all of your tickets win or all lose.
- The number of tickets you buy must be between 0 and 10.
- **Please note:** As in round one, you are completely free to choose how many tickets you want to buy.

Please decide how many tickets you want to buy in round two.

I buy _____ tickets in round two. (**Please note:** The number must be between 0 and 10)

Work-3

Please take some time to answer these questions:

1) Which considerations influenced your decision on how many tickets to buy (especially in round two)?

_____

_____

_____

_____

2) Do you regularly play games of chance (slot machines, lottery, sports bets, poker, etc.)? If so, how often?

yes___ no___

_____

_____

_____

3) Are you familiar with the term “House Money Effect“?

yes___ no___

Thank you for participation!

Instructions for *Robustness Check* treatment:

Robustness Check-1

Thank you for participating in this behavioral economics experiment.
Please answer the following questions about yourself:

I am male ____ female ____

I am _____ years old.

I study _____ in the _____. semester.

The experiment you are participating in consists of a two-stage gambling task.
In the beginning, you are endowed with an amount of 8€ that you can stake in two consecutive lotteries. You can keep the money that you have at the end of round two.

The characteristics of the lotteries are as follows:
- You can use your 8€ to buy lottery tickets in both of the two rounds.
- One lottery ticket costs 0.40€ and wins or loses with a probability of 50%.
- If your tickets win, you will receive 1€ per ticket.
- If your tickets lose, they are blanks and you do not receive any money for them.
- The outcomes of all tickets you buy are perfectly positively correlated, i.e., either all tickets win or all tickets lose.
- You are not informed whether you won or lost after the first round is played. The outcome of round one is announced after your decision for round two.
- The number of tickets you can buy is limited to 10 per round.
- **Please note:** There is no such thing as a right or wrong behavior in this experiment. It is solely up to you, how many tickets you want to buy.

Please decide how many lottery tickets you want to buy in round one.

I buy _____ tickets in round one. (**Please note:** The number must be between 0 and 10)

Robustness Check-2

In the second round of the experiment, you have the possibility to buy lottery tickets again. The conditions are the same as in round one:

- One ticket costs 0.40€
- Your tickets win with a probability of 50% and you receive 1€ per ticket.
- Either all of your tickets win or all lose.
- The number of tickets you buy must be between 0 and 10.
- **Please note:** As in round one, you are completely free to choose how many tickets you want to buy.

Please decide how many tickets you want to buy in round two.

I buy _____ tickets in round two. (**Please note:** The number must be between 0 and 10)

Robustness Check-3

Please take some time to answer these questions:

1) Which considerations influenced your decision on how many tickets to buy (especially in round two)?

_____

_____

_____

_____

2) Do you regularly play games of chance (slot machines, lottery, sports bets, poker, etc.)? If so, how often?

yes___ no___

_____

_____

_____

3) Are you familiar with the term “House Money Effect“?

yes___ no___

Thank you for participation!

Appendix B: Additional statistics

For 2.3.1:

Table 8: Differences in risk attitude between treatments

	Baseline	Time 1	Time 2 / 0	Time 2 / 1
Time 1	$p = 0.0732$*	-		
Time 2 / 0	$p = 0.2596$	$p = 0.4726$	-	
Time 2 / 1	$p = 0.5724$	$p = 0.2290$	$p = 0.2911$	-
Work	$p = 0.0000$***	$p = 0.0254$**	$p = 0.0021$***	$p = 0.0005$***

*Note: two-sided t tests; *=p<0.1, **=p<0.05, ***=p<0.01*

For 2.3.2:

Table 9: Differences in risk attitude between rounds within treatments

Treatment	**Observations**	**Mean / Round 1**	**Mean / Round 2**	**p-value**
All (except Time 2)/winners	n=65	4.43 (1.37)	4.37 (2.93)	0.8631
All (except Time 2)/losers	n= 66	4.47 (1.56)	4.70 (3.18)	0.5431
Baseline /winners	n=22	4.73 (1.28)	5.18 (2.58)	0.4238
Baseline /losers	n=21	4.81 (1.33)	5.57 (3.08)	0.2916
Time 1/winners	n=22	4.41 (1.40)	3.86 (3.11)	0.3939
Time 1/losers	n=22	4.23 (1.66)	4.45 (2.96)	0.6156
Time 2/winners	n=19	4.68 (1.49)	2.84 (1.57)	0.0004***
Time 2/losers	n=23	5.35 (1.61)	4.26 (2.47)	0.0111**
Work/winners	n=21	4.14 (1.42)	4.05 (3.04)	0.8888
Work/losers	n=23	4.39 (1.64)	4.13 (3.43)	0.7275

*Note: two-sided t tests; *=p<0.1, **=p<0.05, ***=p<0.01*

Table 10: Behavioral changes between treatments conditional on first outcome

	Baseline	Time 1	Time 2
Time 1	winners: p =0.2714 losers: p =0.4533		
Time 2	winners***: p =0.0030 losers**: p =0.0437	winners: p =0.1309 losers*: p =0.0998	
Work	winners: p =0.3936 losers: p =0.2749	winners: p =0.7481 losers: p =0.4038	winners**: p =0.0491 losers: p =0.6629

*Note: Mann-Whitney U tests; *=p<0.1, **=p<0.05, ***=p<0.01*

3 Moral Judgement in Probabilistic Dilemmas – A Descriptive Analysis

Abstract

This research employs a multitude of probabilistic versions of two iconic variants of the trolley dilemma. In four studies, subjects rated moral permissibility of action in the bystander case – where one can divert a train, letting one person die instead of five – and the footbridge case – where one can sacrifice one person to save the five – with outcome probabilities being changed systematically. Results show that decreasing attractiveness of intervention yields a decreasing perceived moral permissibility of the intervention. Furthermore, a constant ratio of expected outcomes leaves moral permissibility ratings unchanged on aggregate if outcome probabilities are identical, whereas they display the emergence of a common ratio effect in the bystander, but not in the footbridge situation in case of asymmetric probabilities if these are manipulated intrapersonally. Additionally, probability framing does not seem to be of major importance and previous findings that moral permissibility of intervention is denoted higher in the bystander case are confirmed.

3.1 Introduction

The trolley must have already travelled a long way in people's minds.

Since their first appearances in the 1960's and 70's (Foot, 1967; Thomson, 1976), a set of hypothetical moral dilemmas involving a runaway trolley threatening to kill a certain number of people and a variety of possible ways to prevent this, usually entailing the death of at least one other individual, have become workhorses of eliciting moral judgements, studying moral reasoning and validating moral theory. Fundamentally different responses to two of the most iconic variants of this set of questions, namely the "bystander" situation – where one can operate a switch to divert the trolley onto a sidetrack, killing one instead of five – and the "footbridge" scenario – where one can shove an overweight man off a bridge onto the track, stopping the trolley, yet causing his death – have aroused researchers' interest for decades. Obviously, acting yields the same consequences in both scenarios, i.e., five lives are saved at the expense of one person's death, and are therefore equivalent from a strict consequentialist point of view. Furthermore, it seems legitimate to rate one person dying as better (or let us say, less bad) than five people dying if assessed in isolation. Still, it is a robust empirical finding that most individuals condemn shoving the overweight man but consider it permissible to operate the switch (e.g., Hauser et al., 2007; Lanteri et al., 2008; Broeders et al., 2011; Shallow et al., 2011; Ahlenius and Tännsjo, 2012; Cao et al., 2017).

Apparently, two diametrically opposed approaches to moral behavior play a vital role in this regard. On the one hand, consequentialism states that the morality of an action is solely determined by its consequences, regardless of the means, whereas, on the other hand, deontological theories ascribe an innate morality to an action itself. Important ideas of the latter, that can be utilized in the analyses of the responses are the moral distinction between doing harm and allowing harm to happen (Quinn, 1989), as well as the Doctrine of Double Effect, which claims that a negative consequence is morally acceptable if it is merely a foreseen side-effect of reaching a good end rather than an intended means to assure this good outcome (Hauser et al, 2007).[1]

The main goal of the present study is to add external validity to responses in these hypothetical scenarios by systematically adding explicit outcome probabilities to these dilemmas as most actual moral decisions naturally exhibit considerable outcome uncertainty.[2] It is therefore crucial to understand if and how individuals' moral assessments react to probability changes in both kinds of situations, those that are assumed to trigger mostly consequentialist responses (bystander) and those that are presumably subject to stronger deontological restrictions (footbridge).

[1] For a recent overview of proposed explanations, see Cova (2017).

[2] The most prominent example is the debate about how to program the software of autonomously driving cars (e.g., Nyholm and Smids, 2016; Awad et al., 2018).

This study also examines whether well-known behavioral phenomena documented in the literature on risk and uncertainty, e.g., framing effects and common ratio effects, are identifiable in these types of decision-making. A framing effect in this regard means that individuals' answers differ depending on whether identical outcomes are presented in a gain or loss frame. Risk averse behavior has been reported to emerge in the gain domain, while individuals tend to behave risk-lovingly in a loss domain (Kahneman and Tversky, 1979; Tversky and Kahneman, 1981). In the present study, the common ratio effect is defined as any effect brought about by a change in probabilities that preserves the ratio of expected outcomes but changes the absolute difference in outcome probabilities.[3]

Framing effects in the bystander case have been investigated by Petrinovich and O'Neill (1996), who report that a "kill" wording yields less approval with action or inaction compared to a "save" wording in description of the situation. Cao et al. (2017) find a similar result, namely that a negative, i.e., kill frame produces less consequentialist responses compared to a positive, i.e., save frame in the standard bystander scenario, albeit they do not identify this effect in the footbridge dilemma. However, as they increase the number of people dying in case of omission to 15, the effect arises in the footbridge, but no longer in the bystander dilemma. Broeders et al. (2011) add to this by showing that priming subjects with a "do not kill" rule yields less approval with action in the footbridge case compared to priming with a "save lives" rule

The explicit addition of probabilities to moral dilemmas has not gained much attention to date. Most closely related to this study are those by Shenhav and Greene (2010) and Brand and Oaksford (2015). Shenhav and Greene show that moral acceptability of sacrificing one drowning man to save a larger number of people in a rescue-boat setting is negatively related to the probability that the larger group is rescued by another boat anyway. More generally, they find that moral permissibility ratings of letting the single person die is positively related to the expected value of lives lost in case no action is taken. They furthermore find that subjects are marginally less risk averse when a loss frame rather than a gain frame is induced. Brand and Oaksford tested whether permissibility ratings in a set of moral dilemmas including the bystander and footbridge cases react to a change of the probability that action yields the intended consequence. They do find an overall positive relationship. Baron and Leshner (2000) report that approval with a possibly harmful policy is negatively related to the probability that the negative effects will actually occur and increases as a function of the likelihood that the associated benefits are realized.

Recently, attention has been drawn to the fact that individuals who are confronted with hypothetical, and to a certain degree undoubtedly artificial, dilemmas might not be willing to accept

[3] This violation of expected utility's independence axiom is usually defined in a more narrow sense, namely as the effect that the reduction of the probability of an outcome by a constant factor has a greater impact when the outcome was initially certain, as compared to when it was merely probable (Allais, 1953; Tversky and Kahneman, 1986).

the stipulated outcome probabilities (certainty in most cases) for some reason and rather form subjective probabilities in these instances. Shou and Song (2017) and Ryazanov et al. (2018) report that individuals tend to ascribe probabilities to the outcomes that are well below 100% and are linked to moral permissibility ratings in a predicted way, e.g., the perceived probability of the overweight person dying in case his body is used to stop the trolley is higher than the perceived probability of death for the single individual on the sidetrack in case the trolley is diverted, which goes hand in hand with a lower permissibility rating to shove the man on the track. Kortenkamp and Moore (2014) investigate the effects of adding uncertainty to outcomes in moral dilemmas, including the bystander case, by stating that outcomes "might" rather than "will" happen in case of action/omission. They find that outcome uncertainty is associated with a lower rating of moral permissibility of intervention compared to certainty, which cannot be explained by a change of expected values, possibly induced by an asymmetric adjustment of subjective probabilities.

However, to the best of my knowledge, the present research is the first to introduce explicitly and to vary systematically outcome probabilities in the bystander and footbridge dilemma in both, a negative and positive frame to elicit the effects on subjects' moral permissibility ratings in these situations. My framing manipulation is straightforward and on the one hand involves probabilities of survival in the description of the scenarios (positive frame) and on the other probabilities of death (negative frame).

I conducted three studies, which varied with respect to how the probabilities of the various outcomes were modified between questions and ran a fourth study to check whether the results from study 3 carry over to an interpersonal assessment.

The first study aimed at eliciting risk attitude and risk attitude changes as a reaction to a variation of the ratio of expected outcomes. Starting with certain deaths of the five in case of omission and certain death of the one in case of action, the probability of the five dying in case of omission is reduced gradually, while the death of the single individual in case of intervention remains certain. Individuals are influenced by this modification in a predictable way, i.e., they reduce their moral permissibility rating as the attractiveness of intervention declines.

The second study again departs from certain deaths and reduces probability outcomes simultaneously and identically, thereby fixing the ratio of expected outcomes, which leaves moral permissibility ratings unchanged on aggregate.

Study 3 also manipulates probabilities in a way that preserves the ratio of expected deaths, albeit throughout denoting the death of the one as twice as likely as the death of the five. This approach results in what I regard the most remarkable finding of this paper, namely the emergence of a common ratio effect in the bystander, but not in the footbridge case.

Study 4 shows that this is not replicated if probabilities are manipulated between subjects, with a common ratio effect emerging in both dilemma types.

3.2 Method and materials

Participants in studies 1-3 were undergraduate students enrolled in an introductory economics class at Kiel University, Germany. None of them took part in more than one of the studies. They were approached right at the beginning of the tutorials accompanying the lecture and asked to participate in a short survey study. Subsequently, they were randomly assigned to one of four conditions varying with respect to dilemma type and frame (*bystander/positive, bystander/negative, footbridge/positive and footbridge/negative*). I did not vary dilemma types within subjects to avoid order effects that previous studies have documented (Petrinovich and O'Neill, 1996; Lanteri et al., 2008).

Additionally, subjects were only presented scenarios in one frame to prevent confusion and not to make salient that framing of presentation leaves expected outcomes unchanged. They read a set of six (study 1) to seven (study 2 and 3) dilemma descriptions and were instructed to indicate on a 6-point Likert-scale how morally permissible they think acting (operating the switch, shoving the person) is (0=morally completely impermissible, 5=morally completely permissible). I employed a Likert-scale instead of a binary choice to better assess minor changes in moral permissibility ratings resulting from gradual modifications of outcome probabilities. Additionally, I excluded the neutral position to force subjects to take a stand on the moral assessment, as I would have considered choosing this option similar to a statement that the subject does not want to answer the question. After all, it seems disputable to me whether an action can actually be neither morally permissible nor impermissible, so I decided to avoid any possible discussion in this regard.

I presented dilemmas on separate sheets of paper in complete random order and briefed subjects to give their responses to the dilemmas one after another and not to turn back pages. They were furthermore informed that no such thing as "right" or "wrong" answers existed and that the collected data was completely anonymous. At the end, subjects provided demographic information and reported their self-assessed general risk attitude on an 11-point-scale, a question borrowed from the GSOEP[4]. Almost all participants finished the questionnaire within 10 minutes. All materials were in German.

The dilemma descriptions themselves were designed to be as neutral and short as possible. I specifically refrained from explicitly stating that diverting the trolley saves the five people whose lives are at risk in the default (or kills the one person), or that shoving the overweight person off the bridge to stop the trolley does so respectively. After all, such information should be accompanied by statements that the single person on the sidetrack or on the bridge respectively is saved in case of

[4] The German Socio-Economic Panel is a large-scale longitudinal panel dataset gathered by the DIW, Berlin.

omission (or that omission kills the five people) (Kusev et al., 2016), which, together with information about probabilities, would have overloaded the description unnecessarily in my view. I am confident that all possible outcomes of action and omission are unambiguously clear in the representations of the dilemmas. A comprehensive overview of all dilemma representations employed in this research is provided in the Appendix.

3.3 Study 1 (n=200)

The goal of the first study was to elicit how a variation of the ratio of expected outcomes brought about by probability alteration[5] influences moral judgements and whether this differs between dilemma types and/or decision frames. Note that the analysis of probability alteration is within subjects, whereas dilemma type and frame differ between subjects.

200 participants were confronted with a set of six dilemmas each (50 in each of the four conditions). Of those, 48 participants completed the survey correctly[6] in bystander/positive (46% female; mean age: 21.3, s.d. 2.5), 45 in bystander/negative (62% female; mean age: 22, s.d. 4.3), 46 in footbridge/positive (59% female; mean age: 20.6, s.d. 2.3) and 47 in footbridge/negative (53% female; mean age: 21.7, s.d. 2.8).

Table 1 shows the probabilities involved in the six questions in both frames and the respective ratio of expected outcomes.

Table 1: Properties of dilemmas in study 1

	Positive frame	Negative frame	
Dilemma	**Probability of survival of the five (omission)/ Probability of survival of the one (action)**	**Probability of death of the five (omission)/ Probability of death of the one (action)**	**Outcome ratio (Expected deaths omission/action)**
1.1	0% / 0%	100% / 100%	5
1.2	20% / 0%	80% / 100%	4
1.3	40% / 0%	60% / 100%	3
1.4	60% / 0%	40% / 100%	2
1.5	80% / 0%	20% / 100%	1
1.6	90% / 0%	10% / 100%	0.5

Note that expected outcomes are identical across frames, which therefore only differ with respect to presentation of the relevant information. Further note that the consequentialist choice for a risk

[5] Another way to change (expected) outcome ratios would of course be to vary the number of lives at stake instead of probabilities as done by Cao et al. (2017), Shallow et al. (2011), Rai and Holyoak (2010) and Nakamura (2012), or to vary both (Shenhav and Greene, 2010).

[6] In all studies, I removed all subjects from the analysis, who did not provide answers to all questions or gave ambiguous answers (e.g., by making a cross between values on the Likert Scale).

neutral subject would be to act in dilemmas 1.1 to 1.4, while she should be indifferent between action and omission in 1.5 and choose omission in 1.6.

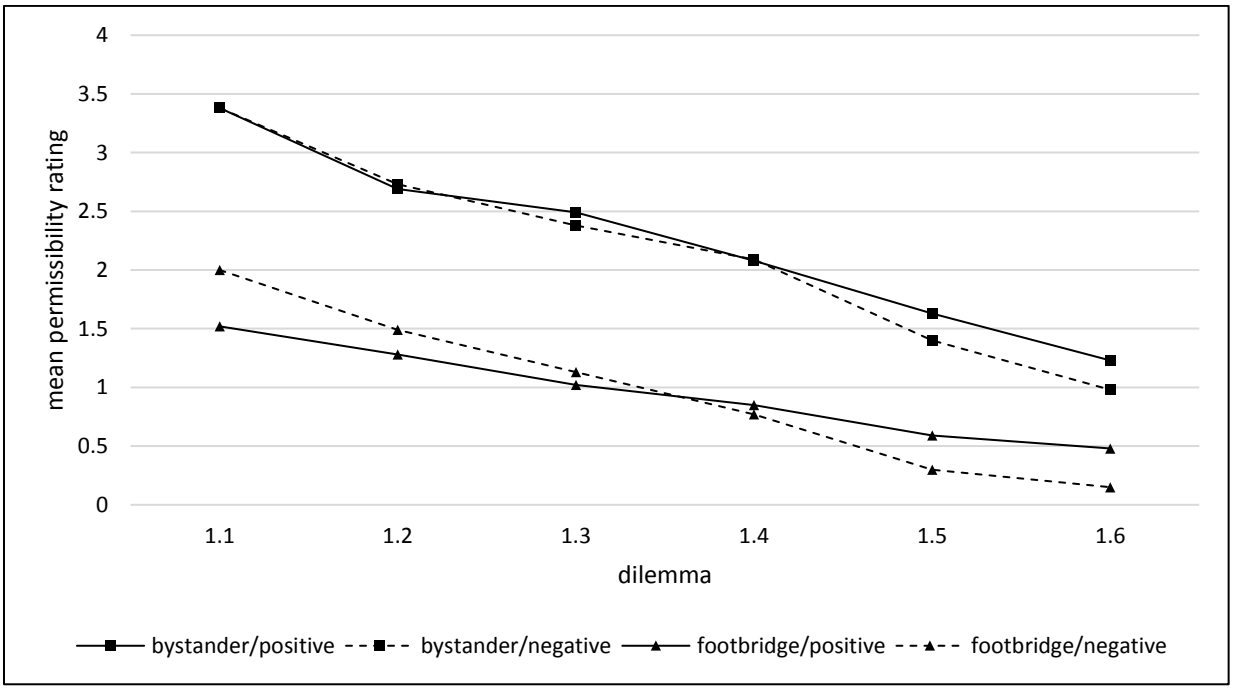
3.3.1 Results

Table 2 depicts the mean moral permissibility ratings for dilemmas 1.1 to 1.6 in all four conditions (standard deviations in parentheses) and figure 1 shows a graphical representation of the results.

Table 2: Descriptive statistics study 1

Dilemma	Frame	bystander	footbridge
1.1	Positive	3.38 (1.59)	1.52 (1.86)
	Negative	3.38 (1.66)	2.00 (1.71)
1.2	Positive	2.69 (1.46)	1.28 (1.52)
	Negative	2.73 (1.53)	1.49 (1.50)
1.3	Positive	2.49 (1.35)	1.02 (1.14)
	Negative	2.38 (1.28)	1.13 (1.19)
1.4	Positive	2.08 (1.40)	0.85 (1.15)
	Negative	2.09 (1.24)	0.77 (1.00)
1.5	Positive	1.63 (1.31)	0.59 (1.02)
	Negative	1.40 (1.14)	0.30 (0.62)
1.6	Positive	1.23 (1.40)	0.48 (0.91)
	Negative	0.98 (1.14)	0.15 (0.36)

Figure 1: Mean permissibility ratings study 1



I used a repeated measures ANOVA with two between subject factors (type and frame) and one within subject factor (probability). Dilemma type has a significant main effect ($F_{1,910} = 66.86, p < 0.01$), while the effect of framing is found insignificant, which also holds for the interaction of type and frame. The employed probability alteration yields a significant effect ($F_{5,910} = 110.77, p < 0.01$)[7]. It is confirmed that this effect is present in all four variants by use of repeated measures ANOVAs in all of the conditions separately ($bystander/positive$: $F_{5,235} = 34.16, p < 0.01$; $bystander/negative$: $F_{5,220} = 36.08, p < 0.01$; $footbridge/positive$: $F_{5,225} = 9.54, p < 0.01$; $footbridge/negative$: $F_{5,230} = 38.75, p < 0.01$).

Using a Bonferroni correction, the results of post-hoc pairwise comparisons of means in all conditions are presented in Table 3.

Table 3: Significance levels of pairwise comparisons in study 1

Dilemma	1.2	1.3	1.4	1.5	1.6
1.1	bp:*** bn:** fp:/ fn:**	bp:*** bn:*** fp:/ fn:***	bp:*** bn:*** fp:*** fn:***	bp:*** bn:*** fp:*** fn:***	bp:*** bn:*** fp:*** fn:***
1.2		bp:/ bn:/ fp:/ fn:/	bp:** bn:** fp:/ fn:***	bp:*** bn:*** fp:*** fn:***	bp:*** bn:*** fp:*** fn:***
1.3			bp:/ bn:/ fp:/ fn:/	bp:*** bn:*** fp:/ fn:***	bp:*** bn:*** fp:* fn:***
1.4				bp:/ bn:** fp:/ fn:*	bp:*** bn:*** fp:/ fn:***
1.5					bp:/ bn:/ fp:/ fn:/

*Note: bystander/positive = bp, bystander/negative = bn, footbridge/positive = fp, footbridge/negative =fn, Bonferroni correction. /=insignificance, *=p<0.10, **=p<0.05, ***=p<0.01.*

Hence, I conclude that we observe a clear downward trend in mean permissibility assessments as the ratio of expected outcomes declines.

Additionally, a significant interaction effect between probability and type is observed ($F_{5,910} = 4.59, p < 0.01$), as well as a significant interaction between probability and frame ($F_{5,910} = 4.59, p < 0.01$, the level of significance declines even by use of the less conservative Huynh-Feldt correction, $p = 0.06$). The first observation is best explained by a floor effect of moral assessment in the bystander cases. The latter presumably roots in different reactions to probability alterations in the footbridge cases. This reading is supported by repeated measures ANOVAs investigating the effect of framing and probabilities for the two types separately. The interaction of probability and frame shows insignificance for the bystander case, but is found significant for the footbridge case ($F_{5,455} =$

[7] If not stated otherwise, the reported significance levels are also met when employing a conservative corrective measure (Greenhouse-Geisser correction) to counter detected lack of sphericity.

$3.21, p < 0.01$, Greenhouse-Geisser correction: $p = 0.041$), while the main effect of framing is found insignificant in both variants.

Furthermore, I performed pairwise comparisons of the proportions of individuals who reported at least two distinct ratings across dilemmas between all conditions while excluding all those subjects who reported complete moral impermissibility for all six dilemmas from the analysis (1 in *bystander/positive*, 2 in *bystander/negative*, 15 in *footbridge/positive* and 12 in *footbridge/negative*). 43 of 47 changed at least once in *bystander/positive*, 42 of 43 in *bystander/negative*, 28 of 31 in *footbridge/positive* and 32 of 35 in *footbridge/negative*. None of the differences in proportions is significant.

Additionally, I ran Tobit regressions to assess the influence of self-reported general risk attitude on moral permissibility ratings in each of the dilemmas.[8] Results show a marginally negative effect in *bystander/negative, 1.3* (p<0.1), significant positive effects in *footbridge/positive, 1.2, 1.4, 1.6* (p<0.05) and insignificance otherwise.

3.3.2 Discussion

The results confirm previous findings that acting in the footbridge dilemma is deemed less permissible than in the bystander dilemma. They add that this difference persists when outcome probabilities are changed gradually as presented.

It is also evident that changing the ratio of expected outcomes in this way, i.e., reducing the relative attractiveness of intervention while leaving the outcome of intervention unchanged, results in the behavioral changes that would be expected whenever the importance of consequences is non-negligible. Most notably, this holds for both the bystander and footbridge cases, which leads me to conclude here that consequentialist deliberations are prominently present in both of these variants.

Different frames, on the other hand, do not seem to be of notable importance here, at least not in the way I employed them. One lesson suggested by this may be that subjects tend to be resistant to different presentations of identical outcomes in these contexts.

Risk attitude does not seem to play a vital role in determining subjects' answers, although it is striking enough that the rare significant effects tend to be positive. A stronger risk aversion (as indicated by a lower Likert-score of risk attitude) would be expected to result in a lower disposition to leave the five lives at risk and should therefore result in a stronger inclination to opt for the certain outcome, i.e., sacrificing one life with certainty. I will return to this issue later.

[8] Please note that results are not independent between dilemmas within one condition as the same subjects stated their moral assessments for each specification. The same applies to studies 2 and 3.

3.4 Study 2 (n=200)

This study is designed to test the influence of probabilities on moral judgements without altering the expected outcome ratio, i.e., without making one alternative more attractive with respect to expected outcomes. Again, the analysis of probability alteration is within subjects and dilemma type and frame differ between subjects.

200 participants were confronted with a set of seven dilemmas each (50 in each of the four conditions). Of those, 45 participants completed the survey correctly in bystander/positive (36% female; mean age: 20.8, s.d. 2.1), 45 in bystander/negative (56% female; mean age: 20.7, s.d. 1.9), 45 in footbridge/positive (51% female; mean age: 20.8, s.d. 1.8) and 45 in footbridge/negative (60% female; mean age: 20.9, s.d. 2.1).

Table 4 shows the probabilities involved in the seven questions in both frames and the respective (constant) ratio of expected outcomes.

Table 4: Properties of dilemmas in study 2

	Positive frame	Negative frame	
Dilemma	**Probability of survival of the five (omission)/ Probability of survival of the one (action)**	**Probability of death of the five (omission)/ Probability of death of the one (action)**	**Outcome ratio (Expected deaths omission/action)**
2.1	0% / 0%	100% / 100%	5
2.2	1% / 1%	99% / 99%	5
2.3	20% / 20%	80% / 80%	5
2.4	40% / 40%	60% / 60%	5
2.5	60% / 60%	40% / 40%	5
2.6	80% / 80%	20% / 20%	5
2.7	99% / 99%	1% / 1%	5

Again, expected consequences are identical across frames. Note that the consequentialist choice would be to act in all specifications. Furthermore, any set of choices satisfying the independence axiom of expected utility theory should be characterized by the same assessment in all dilemmas.

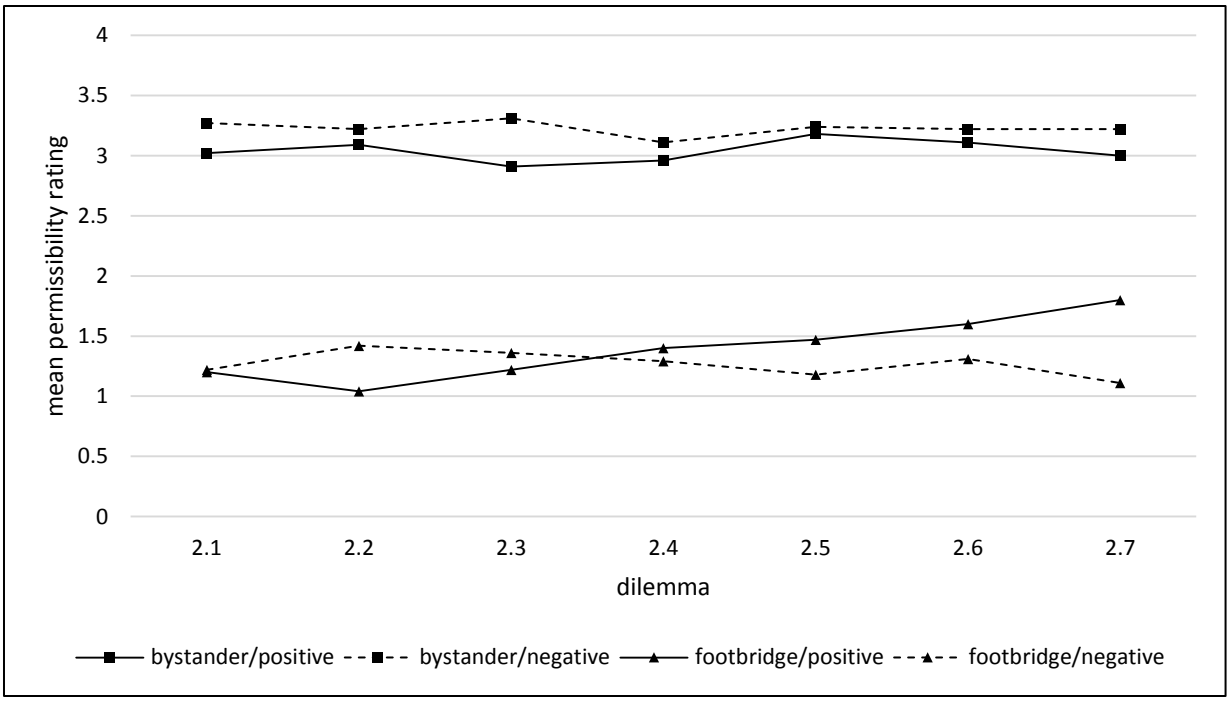
3.4.1 Results

Table 5 depicts the mean moral permissibility ratings for dilemmas 2.1 to 2.7 in all four conditions (standard deviations in parentheses) and figure 2 shows a graphical representation of the results.

Table 5: Descriptive statistics study 2

Dilemma	Frame	bystander	footbridge
2.1	Positive	3.02 (1.54)	1.2 (1.63)
	Negative	3.27 (1.66)	1.22 (1.61)
2.2	Positive	3.09 (1.56)	1.04 (1.41)
	Negative	3.22 (1.54)	1.42 (1.57)
2.3	Positive	2.91 (1.50)	1.22 (1.40)
	Negative	3.31 (1.20)	1.36 (1.49)
2.4	Positive	2.96 (1.36)	1.40 (1.39)
	Negative	3.11 (1.27)	1.29 (1.46)
2.5	Positive	3.18 (1.27)	1.47 (1.46)
	Negative	3.24 (1.17)	1.18 (1.39)
2.6	Positive	3.11 (1.51)	1.60 (1.62)
	Negative	3.22 (1.54)	1.31 (1.50)
2.7	Positive	3.00 (1.86)	1.80 (1.98)
	Negative	3.22 (1.70)	1.11 (1.47)

Figure 2: Mean permissibility ratings study 2



I used a repeated measures ANOVA with two between subject factors (type and frame) and one within subject factor (probability). Dilemma type has a significant main effect ($F_{1,1056} = 89.79, p < 0.01$), while the effect of framing is found insignificant, which also holds for the interaction of type and frame.

The main effect of the employed probability alteration is insignificant as well as the interaction between probability and type, but the interaction between probability and frame shows some significance ($F_{6,1056} = 2.16, p < 0.05$, Huynh-Feldt correction: $p = 0.099$; Greenhouse-Geisser correction: $p = 0.101$). Running repeated measures ANOVAs investigating the effect of framing and probabilities for the two types separately reveals that this effect is driven by significant interaction between probability and frame in the footbridge case ($F_{6,528} = 4.30, p < 0.01$), while it is insignificant in the bystander case. However, main effects of framing and probability alteration are found insignificant in these analyses. Checking the effect of probability alterations with repeated measures ANOVAs in each of the four conditions separately shows insignificant effects except for f*ootbridge/positive* ($F_{6,264} = 4.30, p < 0.01$).

Using a Bonferroni correction, the results of post-hoc pairwise comparisons of means in this condition are presented in Table 6.

Table 6: Significance levels of pairwise comparisons in study 2

Dilemma	2.2	2.3	2.4	2.5	2.6	2.7
2.1	fp:/	fp:/	fp:/	fp:/	fp:/	fp:**
2.2		fp:/	fp:/	fp:/	fp:**	fp:***
2.3			fp:/	fp:/	fp:/	fp:**
2.4				fp:/	fp:/	fp:/
2.5					fp:/	fp:/
2.6						fp:/

*Note: footbridge/positive = fp, Bonferroni correction. /=insignificance, *=p<0.10, **=p<0.05, ***=p<0.01*

Results show a weak tendency for an increased moral permissibility rating as survival probabilities increase in this condition.

Again, a pairwise comparison of the proportions of individuals who reported at least two distinct ratings across dilemmas was conducted between all conditions while excluding all those subjects who reported complete moral impermissibility for all six dilemmas from the analysis (none in *bystander/positive* and *bystander/negative*, 11 in *footbridge/positive* and 15 in *footbridge/negative*). 39 of 45 changed at least once in *bystander/positive*, 32 of 45 in *bystander/negative*, 30 of 34 in *footbridge/positive* and 23 of 30 in *footbridge/negative*. Results show that the proportion is marginally significantly (p<0.10) lower in *bystander/negative* compared to *bystander/positive* and *footbridge/positive*, but no significance was detected in the remaining comparisons.

In addition, Tobit regressions were run to assess the influence of self-reported general risk attitude on moral permissibility ratings in each of the dilemmas. I find a marginally significant positive effect in *bystander/positive, 2.7* (p<0.1), a significant positive effect in *bystander/negative, 2.4* (p<0.05), marginally significant positive effects in *footbridge/positive, 2.3, 2.4* (p<0.1) and significant

positive effects in *footbridge/positive, 2.2, 2.7* ($p<0.05$). Effects in *bystander/negative* are positive and highly significant ($p<0.01$), with the only exception being dilemma 2.7, where a lower level of significance was detected ($p<0.05$).

3.4.2 Discussion

Again, the influence of dilemma type replicates previous findings throughout.

It is also straightforward to conclude that this manipulation of probabilities that preserves the ratio of expected outcomes does not change moral permissibility assessments in a predictable way. I regard this result as additional evidence that ratios of expected outcomes, and thus consequentialist properties, are an important determinant of moral permissibility assessments. Due to effects being weak and exceptional, I would refrain from interpreting the results in *footbridge/negative* as a challenge to this general finding.

However, most individuals do not consistently stick to their moral assessment across dilemmas, but change it at least once. Hence, caution is in order when applying this finding at the individual level, yet it seems to hold on aggregate.

Similar to study 1, the influence of framing does not seem to be of major importance, with the same implications being applicable.

Again, risk attitude appears to be positively linked to perceived moral permissibility of putting the one person at risk rather than leaving five lives at risk. As stated before, this peculiar finding will be tackled later on.

3.5 Study 3 (n=200)

This study is designed to render the constancy of ratios of expected outcomes less salient and allow for an asymmetrical alteration of probabilities. It also serves to control for common ratio effects in moral decision-making as defined previously. As before, the analysis of probability alteration is within subjects, whereas dilemma type and frame differ between subjects.

200 participants were confronted with a set of seven dilemmas each (50 in each of the four conditions). Of those, 46 participants completed the survey correctly in bystander/positive (59% female; mean age: 20.6, s.d. 1.9), 45 in bystander/negative (56% female; mean age: 21.1, s.d. 2.5), 45 in footbridge/positive (56% female; mean age: 21.1, s.d. 3.0) and 48 in footbridge/negative (67% female; mean age: 20.9, s.d. 2.1).

Table 7 shows the probabilities involved in the seven questions in both frames and the respective (constant) ratio of expected outcomes.

Table 7: Properties of dilemmas in study 3

	Positive frame	Negative frame	
Dilemma	**Probability of survival of the five (omission)/ Probability of survival of the one (action)**	**Probability of death of the five (omission)/ Probability of death of the one (action)**	**Outcome ratio (Expected deaths omission/action)**
3.1	50% / 0%	50% / 100%	2.5
3.2	51% / 2%	49% / 98%	2.5
3.3	60% / 20%	40% / 80%	2.5
3.4	70% / 40%	30% / 60%	2.5
3.5	80% / 60%	20% / 40%	2.5
3.6	90% / 80%	10% / 20%	2.5
3.7	99% / 98%	1% / 2%	2.5

Expected outcomes are identical across frames. Note that the consequentialist choice of a risk neutral individual would be to act in all specifications. Furthermore, any set of choices satisfying the independence axiom of expected utility theory should again be characterized by the same assessment in all dilemmas.

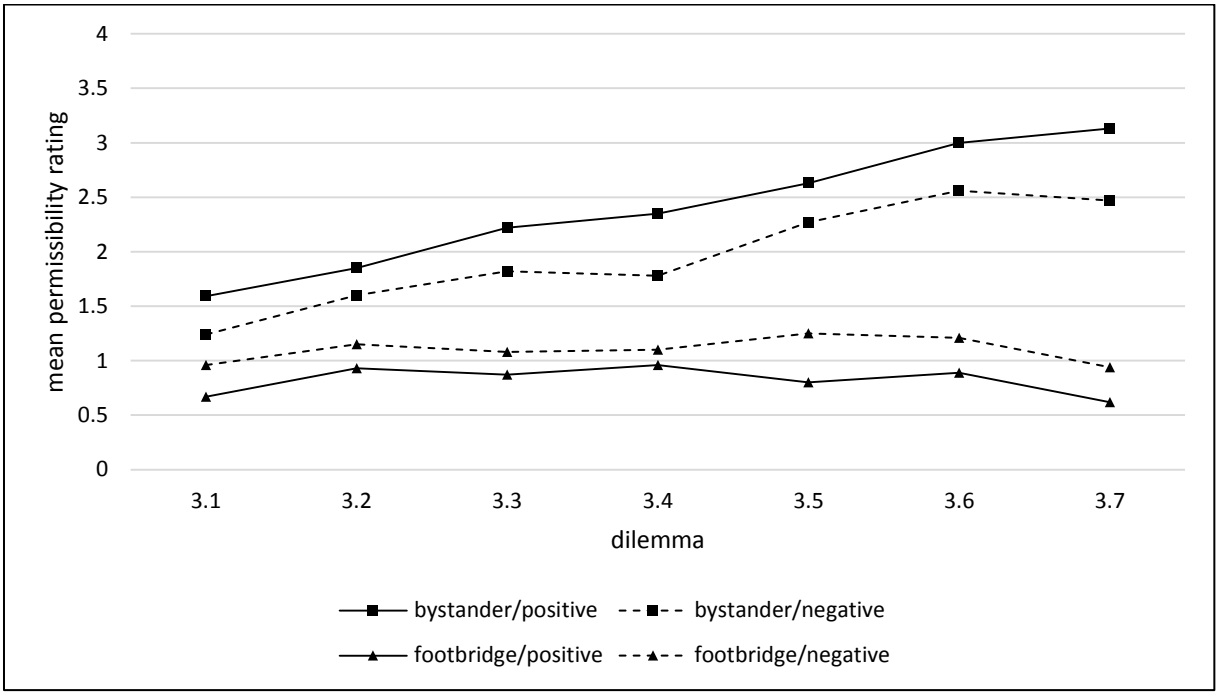
3.5.1 Results

Table 8 depicts the mean moral permissibility ratings for dilemmas 3.1 to 3.7 in all four conditions (standard deviations in parentheses) and figure 3 shows a graphical representation of the results.

Table 8: Descriptive statistics study 3

Dilemma	Frame	bystander	footbridge
3.1	Positive	1.59 (1.36)	0.67 (1.21)
	Negative	1.24 (1.19)	0.96 (1.46)
3.2	Positive	1.85 (1.43)	0.93 (1.39)
	Negative	1.60 (1.48)	1.15 (1.53)
3.3	Positive	2.22 (1.38)	0.87 (1.25)
	Negative	1.82 (1.48)	1.08 (1.44)
3.4	Positive	2.35 (1.20)	0.96 (1.19)
	Negative	1.78 (1.29)	1.10 (1.43)
3.5	Positive	2.63 (1.22)	0.80 (1.06)
	Negative	2.27 (1.44)	1.25 (1.51)
3.6	Positive	3.00 (1.46)	0.89 (1.37)
	Negative	2.56 (1.55)	1.21 (1.56)
3.7	Positive	3.13 (1.77)	0.62 (1.17)
	Negative	2.47 (1.89)	0.94 (1.36)

Figure 3: Mean permissibility ratings study 3



As before, I used a repeated measures ANOVA with two between subject factors (type and frame) and one within subject factor (probability). Dilemma type has a significant main effect ($F_{1,1080} = 48.29, p < 0.01$), while the main effect of framing is found insignificant. The interaction of these two is significant ($F_{1,1080} = 4.11, p < 0.05$). Running repeated measures ANOVAs investigating the effect of framing and probabilities for the two types separately reveals that the main effect of framing is marginally significant in the bystander dilemma ($F_{1,534} = 4.11, p < 0.1$), but insignificant in the footbridge cases. The main effect of probability alteration is significant ($F_{6,1080} = 20.58, p < 0.01$), which also holds for the interaction of probability and type ($F_{6,1080} = 19.43, p < 0.01$), while the interaction of probability and frame is found insignificant. The repeated measures ANOVAs run for the two types separately show significant main effects of probability alteration for both cases (bystander: $F_{6,534} = 26.98, p < 0.01$; footbridge: $F_{6,546} = 3.01, p < 0.01$, Huynh-Feldt correction: $p = 0.02$).

I next checked the effects of probability alteration using a repeated measures ANOVA in each of the four conditions separately. Results show highly significant effects in *bystander/positive* ($F_{6,270} = 15.98, p < 0.01$) and *bystander/negative* ($F_{6,264} = 11.46, p < 0.01$). The effect is insignificant in *footbridge/negative* ($F_{6,282} = 1.74, p = 0.11$), but marginally significant in *footbridge/positive* ($F_{6,264} = 1.80, p < 0.1$). However, even if a less conservative correction for the observed lack of sphericity is used here (Hynh-Feldt correction), the effect is found insignificant ($F_{6,282} = 1.80, p = 0.15$) and the condition therefore not considered in the following investigation.

Using a Bonferroni correction, the results of post-hoc pairwise comparisons of means in the two *bystander* conditions are presented in Table 9.

Table 9: Significance levels of pairwise comparisons in study 3

Dilemma	3.2	3.3	3.4	3.5	3.6	3.7
3.1	bp:/ bn:/	bp:** bn:*	bp:*** bn:/	bp:*** bn:***	bp:*** bn:***	bp:*** bn:***
3.2		bp:/ bn:/	bp:/ bn:/	bp:*** bn:**	bp:*** bn:***	bp:*** bn:***
3.3			bp:/ bn:/	bp:/ bn:/	bp:*** bn:***	bp:*** bn:**
3.4				bp:/ bn:/	bp:** bn:***	bp:*** bn:**
3.5					bp:/ bn:/	bp:/ bn:/
3.6						bp:/ bn:/

*Note: bystander/positive = bp, bystander/negative = bn, Bonferroni correction. /=insignificance, *=p<0.10, **=p<0.05, ***=p<0.01.*

Obviously, there is a strong tendency for moral permissibility to increase as probabilities of survival (death) are increased (decreased) asymmetrically with the ratio of expected outcomes remaining unchanged.

As in the previous studies, I performed pairwise comparisons of the proportions of individuals who reported at least two distinct ratings across dilemmas between all conditions while excluding all those subjects who reported complete moral impermissibility for all six dilemmas from the analysis (3 in *bystander/positive*, 5 in *bystander/negative*, 19 in *footbridge/positive* and 18 in *footbridge/negative*). 40 of 43 changed at least once in *bystander/positive*, 36 of 40 in *bystander/negative*, 24 of 26 in *footbridge/positive* and 25 of 30 in *footbridge/negative*. None of the differences in proportions is significant.

Results of Tobit regressions of self-reported risk attitude on moral permissibility ratings show either insignificant or positive effects. These positive associations are observed for *bystander/positive, 3.1* ($p<0.05$), *3.5* ($p<0.1$), *3.7* ($p<0.05$), bystander/negative, *3.4* ($p<0.1$), *3.5* ($p<0.01$), *3.6* ($p<0.1$), *3.7* ($p<0.01$) and *footbridge/negative, 3.1* ($p<0.1$), *3.2, 3.3, 3.4, 3.5, 3.6* ($p<0.05$).

3.5.2 Discussion

Results show again that moral permissibility of intervention is generally perceived higher in the bystander case than in the footbridge dilemma, which does not come as a surprise at this point.

I do again observe that framing is not of substantial importance and that a higher reported risk attitude tends to increase individual's inclination to deem an intervention as morally permissible. In both footbridge conditions, the unchanged ratio of expected outcomes goes hand in hand with a lack of statistically significant differences in assessment of morality on aggregate. However, one should

again be cautious to apply this to the individual level, as most subjects reported at least two distinct ratings across dilemmas.

Clearly, the most striking result in this study is that moral permissibility ratings increase substantially in both bystander specifications as probabilities of death decline, although the ratio of expected outcomes stays constant. I regard this as a clear indication that a common ratio effect is present in these conditions. The explanation for the emergence of this effect might include that individuals put more weight on the absolute difference in outcome probabilities rather than on the constancy of expected outcome ratios. This effect is probably even amplified if the death of the single individual is denoted highly probable or even certain when compared to the much less probable death of the five. A possible explanation for the same rationale not applying to the footbridge cases is provided by Greene's dual-process theory (Greene et al., 2001; Greene et al., 2004; Greene, 2007), which is sketched in the last section of this paper.

3.6 Study 4 (n=420)

To check whether the interaction found in study 3 prevails if probabilities are manipulated interpersonally instead, I conducted an Online-Survey that was set up via oTree (Chen et al., 2016).
420 subjects judged moral permissibility for only one dilemma each. The dilemmas employed in this study are 3.1 (*bystander/positive* and *footbridge/positive*) and 3.7 (*bystander/positive* and *footbridge/positive*). Table 7 depicts the mean moral permissibility ratings (standard deviations in parentheses).

Table 10: Descriptive statistics study 4

Dilemma	Frame	bystander	footbridge
3.1	Positive	2.40 (1.47), n=108	1.12 (1.43), n=114
3.7	Positive	3.10 (1.65), n=101	1.59 (1.80), n=97

A two-way factorial ANOVA renders the main effect of dilemma type significant ($F_{1,416} = 81.02, p < 0.001$), which is also true for the effect of probability manipulation ($F_{1,416} = 14.18, p < 0.001$), with the interaction of these two being found insignificant ($F_{1,416} = 0.58, p = 0.45$). Separately employing ANOVAs for each of the types shows the effectiveness of probability alteration in both, bystander dilemma ($F_{1,207} = 10.55, p < 0.01$) and footbridge dilemma ($F_{1,209} = 4.38, p < 0.05$), while it is slightly less pronounced in the latter. Hence, a common ratio effect emerges for both dilemma types if comparing permissibility ratings interpersonally.

3.7 General discussion

Let us first take a closer look at the lack of considerable framing effects in all of the three studies, as this is at odds with previous findings mentioned earlier. A reason for subjects allegedly paying more attention to actual outcomes might be that the consequences of their actions are to some degree decoupled from the actions themselves in the descriptions of the dilemmas at hand. While all possible outcomes are unambiguously clear in my view, descriptions did not involve explicitly that action or omission "saves", "kills" or "sacrifices" the respective number of individuals as in previous studies or "puts/leaves them at risk" (refer to the Appendix for the exact wording of the dilemmas). Less subtle differences might spawn the effects found previously.

Some general words of caution are in order when interpreting the results of all three studies, at least with respect to the intrapersonal analysis. Subjects answered multiple questions each and are therefore naturally susceptible to anchor effects (Tversky and Kahneman, 1974) and, more generally, experimenter demand effects (Zizzo, 2010), which may result in substantial order effects. Individuals in study 1 might take their assessment of moral permissibility in the dilemma they encounter first as a starting point and arrange their following assessments around this in a way that they think is appropriate or expected. [9] Remember that different ratios of expected outcome are salient in study 1. However, even if one interprets adjustments of ratings as solely driven by experimenter demand effects[10], it is obvious that these adjustments occur in a predictable direction, therefore indicating that subjects do at least anticipate a consequentialist assessment of morality as discussed before. Nevertheless, as the observed changes are unambiguous and strong, it is reasonable to assume that they are real effects rather than cognitive mistakes.

The same effects might be at work in study 2, although it is worth noting that the direction of effects is ambiguous here. Individuals might feel inclined or expected to change their assessments between dilemmas, while not being fully aware about the expected direction, as ratios of expected outcomes are obviously constant. This might be one reason why individuals change their assessment as discussed, but in an unpredictable way on aggregate.

Study 3 stands out in this regard, as individuals in the bystander conditions do react to changes of outcome probabilities in a predictable way, but do not in the footbridge conditions, yet a substantial fraction of subjects changes assessments at least once between dilemmas in all conditions.

Although it is a scattered effect, the positive connection between self-reported risk attitude and reported ratings of morality that appears some of the time is surprising and needs closer inspection. First, it is noteworthy in this regard that asking for moral permissibility ratings and having individuals state whether they would actually intervene in the respective dilemma themselves is not equivalent

[9] This possible reading is the main reason why I confronted subjects with the dilemmas in random order.
[10] A view that would highly overestimate the effect in my view.

per se. Therefore, answers driven by individual risk attitude to the latter question do not necessarily transfer to the former, which in turn does not exclude the possibility that people would have answered in line with theory if asked for their preferred action, i.e., would have shown a negative relationship between risk attitude and inclination to intervene in study 1. Some further deliberations may shed some light on the findings, although they should be treated with caution due to their speculative nature. First, as all subjects report their risk attitude after having rated the moral permissibility in the respective dilemmas, some reverse causality may be at work. Namely, a higher rating of "action" as permissible may convince individuals that they are in fact risk-loving and therefore increase subsequently reported risk-attitude.

Second, it is quite possible that individuals do pay less attention to the risk associated with the default compared to the risk that is posed on the single individual in case of intervention, thus requiring a higher risk proneness to rate intervention as permissible. After all, the initial risk in the default is not the result of the individual's decision, yet new risk is created in case of taking the action. This reading would explain to a certain degree why risk attitude does not seem to be as influential in study 1, where no risk is associated with intervention and the single person dies with certainty in all specifications.
Third, moral permissibility ratings and risk attitude may not be causally linked in these types of decisions after all, but possibly driven by some unobserved other determinant(s).

In my opinion, undoubtedly the most interesting and surprising finding in study 3 is the emergence of a common ratio effect in the bystander conditions while it is unobservable in the footbridge cases. Different assessments of subjective outcome probabilities in the two dilemmas cannot possibly account for this finding in my view. Study 1 has shown that individuals are susceptible to variations of probability of the five dying in case of omission. Combined with results from study 2, I conclude that subjects are also willing to accept how the probability of the single person dying in case of intervention is varied in both dilemmas.

Greene's dual-process theory (Greene et al., 2001; Greene et al., 2004; Greene, 2007) provides a consistent explanation of the results. In short, it states that people are driven by negative emotional responses in reaction to the thought of pushing the person off the bridge, but also engage in consequentialist reasoning. If these conflict, more cognitive control is provoked to possibly override the emotional response. This idea is obviously consistent with the observed reduction of permissibility ratings as the attractiveness of intervention objectively declines in both dilemmas. It also explains the difference between dilemma types as the emotional response is relatively weaker in the bystander case. Framing effects turn out negligible as the differences in framing are decoupled from the action itself, and therefore from the emotional response. Under the assumption that higher cognitive control is spawned in the footbridge case, responses are prevented from exhibiting a common ratio effect as individuals engage in more thorough cost-benefit analysis and are not influenced by the difference in

absolute probabilities, as is the case in the bystander dilemma. The fact that a common ratio effect emerges also for the footbridge case in an intrapersonal analysis provides further support for this explanation, as the higher cognitive control that is invoked prevents subjects from answering in a subjectively inconsistent way if confronted with different dilemmas with a constant outcome ratio. This motivation is absent in the interpersonal analysis, which shows that judgements may vary even if the ratio is kept constant.

To conclude, the present paper offers some initial results from enriching well-known moral dilemmas with explicit outcome probabilities. First, the regularly observed discrepancy in moral assessments between these dilemmas persists if outcomes are denoted risky in various ways. Second, assessments in both dilemmas follow some consequentialist reasoning if probabilities are varied in a way that alters the ratio of expected outcomes. Third, moral permissibility ratings are unaffected by symmetric alterations of outcome probabilities on aggregate. Fourth, a common ratio effect is present in the bystander dilemma but not in the footbridge dilemma if probabilities are manipulated intrapersonally. Additionally, probability framing as employed does not yield considerable effects.

Some additional directions for future research might include the investigation of group decisions in probabilistic moral dilemmas, as many real world decisions that are possibly a matter of life and death are not made by one person alone. Once again, programming the software of self-driving cars can count as a textbook example in this regard, as well as does developing guidelines for medical care provision. It also seems promising to connect additional behavioral insights from economic research to judgements in moral dilemmas. It might for example make a difference in the spirit of a sunk cost effect, whether the trolley has or has not already run over and killed a certain number of individuals before the actual decision is to be made. As a final remark, it is of course possible to alter outcome probabilities in a variety of other ways than the ones chosen here. One could for example design dilemmas or modify the trolley dilemmas such that outcomes for both parties are uncertain, while this research only employed dilemmas where the outcome for one party, determined by the choice between intervention and omission, was certain safety and only, possibly, probabilistic for the other.

In light of the present studies' results and the entailed further questions, I am confident that the trolley dilemma will retain its role as an important tool to investigate moral decision-making for quite some time to come.

References

Ahlenius, H. & Tännsjö, T. (2012). Chinese and Westerners Respond Differently to the Trolley Dilemmas. *Journal of Cognition and Culture*, 12, 195-201.

Allais, M. (1953). Le Comportement de l'Homme Rationnel devant le Risque: Critique des Postulats et Axiomes de l'Ecole Americaine. *Econometrica*, Vol. 21, No. 4, pp. 503-546.

Awad, E., Dsouza, S., Kim, R., Schulz, J., Henrich, J., Shariff, A., Bonnefon, J.-F. & Rahwan, I. (2018). The Moral Machine experiment. *Nature*, 563, 59-64.

Baron, J. & Leshner, S. (2000). How Serious are Expressions of Protected Values? *Journal of Experimental Psychology: Applied,* 6(3), 183-194.

Brand, C.M. & Oaksford, M. (2015). The Effect of Probability Anchors on Moral Decision Making. In D. C. Noelle, R. Dale, A. S. Warlaumont, J. Yoshimi, T. Matlock, C. D. Jennings, P. P. & Maglio, (Eds.), *Proceedings of the 37th Annual Meeting of the Cognitive Science Society* (pp. 268-272). Austin, TX: Cognitive Science Society.

Broeders, R., van den Bos, K., Müller, P. A. & Ham, J. (2011). Should I save or should I not kill? How people solve moral dilemmas depends on which rule is most accessible. *Journal of Experimental Social Psychology*, 47, 923-934.

Cao, F., Zhang, J., Song, L., Wang, S., Miao, D. & Peng, J. (2017). Framing Effect in the Trolley Problem and Footbridge Dilemma: Number of Saved Lives Matters. *Psychological Reports*, 120(1), 88-101.

Chen, D.L., Schonger, M. & Wickens, C. (2016). oTree – An open-source platform for laboratory, online, and field experiments. *Journal of Behavioral and Experimental Finance,* 9, 88-97.

Cova, F. (2017). What Happened to the Trolley Problem? *Journal of Indian Council of Philosophical Research*, 34, 543-564.

Foot, P. (1967). The problem of abortion and the doctrine of double effect. *Oxford Review*, 5, 5–15.

Greene, J. D., Sommerville R. B., Nystrom L. E., Darley J. M. & Cohen J. D. (2001). An fMRI Investigation of Emotional Engagement in Moral Judgement. *Science,* 293, 2105-2108.

Greene, J. D., Nystrom L. E., Engell A. D., Darley J. M. & Cohen J. D. (2004). The Neural Bases of Cognitive Conflict and Control in Moral Judgment. *Neuron,* 44, 389-400.

Greene, J. D. (2007). Why are VMPFC patients more utilitarian? A dual-process theory of moral judgment explains. *Trends in Cognitive Sciences, 11*(8), 322-323.

Hauser, M. D., Cushman, F. A., Young, L., Jin, R. & Mikhail, J. M. (2007). A dissociation between moral judgment and justification. *Mind and Language*, 22, 1–21.

Kahneman, D. & Tversky, A. (1979). Prospect Theory: An Analysis of Decision under Risk. *Econometrica*, 47(2), 263-291.

Kortenkamp, K. V. & Moore, C. F. (2014). Ethics Under Uncertainty: The Morality and Appropriateness of Utilitarianism When Outcomes are Uncertain. *The American Journal of Psychology*, 127(3), 367-382.

Kusev, P., van Schaik, P., Alzahrani, S., Lonigro, S. & Purser, H. (2016). Judging the morality of utilitarian actions: How poor utilitarian accessibility makes judges irrational. *Psychonomic Bulletin and Review*, 23, 1961-1967.

Lanteri, A., Chelini, C. & Rizzello, S. (2008). An experimental investigation of emotions and reasoning in the trolley problem. *Journal of Business Ethics*, 83(4), 789-804.

Nakamura, K. (2012). The Footbridge Dilemma Reflects More Utilitarian Thinking Than The Trolley Dilemma: Effect Of Number Of Victims In Moral Dilemmas. *Proceedings of the Annual Meeting of the Cognitive Science Society*, 34(34), 803-808.

Nyholm, S. & Smids, J. (2016). The Ethics of Accident-Algorithms for Self-Driving Cars: An Applied Trolley Problem? *Ethical Theory and Moral Practice*, 19, 1275-1289.

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

Rai, T. S. & Holyoak, K. J. (2010). Moral Principles or Consumer Preferences? Alternative Framings of the Trolley Problem. *Cognitive Science*, 34, 311-321.

Ryazanov, A. A., Knutzen, J., Rickless, S. C., Christenfeld, N. J. S. & Nelkin, D.K. (2018). Intuitive Probabilities and the Limitation of Moral Imagination. *Cognitive Science*, 1-31.

Shallow, C., Iliev, R. & Medin, D. (2011). Trolley problems in context. *Judgment and Decision Making*, 6(7), 593 – 601

Shenhav, A. & Greene, J. D. (2010). Moral Judgements Recruit Domain-General Valuation Mechanisms to Integrate Representations of Probability and Magnitude. *Neuron*, 67, 667-677.

Shou, Y. & Song, F. (2017). Decisions in moral dilemmas: The influence of subjective beliefs in outcome probabilities. *Judgment and Decision Making*, 12(5), 481.

Thomson, J. J. (1976). Killing, letting die, and the trolley problem. *The Monist*, 59(2), 204–217.

Tversky, A. & Kahneman, D. (1974). Judgement under Uncertainty: Heuristics and Biases. *Science*, 185 (4157), 1124-1131.

Tversky, A. & Kahneman, D. (1981). The Framing of Decisions and the Psychology of Choice. *Science*, 211(4481), 453-458.

Tversky, A. & Kahneman, D. (1986). Rational choice and the framing of decisions. *The Journal of Business*, 59(4), 251-278.

Zizzo, D. J. (2010). Experimenter demand effects in economic experiments. *Experimental Economics*, 13(1), 75-98.

Appendix

Experimental instructions and materials (translated from German)

Study 1-3

Subjects received a total of eight (study 1) or nine (study 2 and study 3) pages. The first page and the last page were the same for all subjects. The pages in between contained one dilemma each in random order that differed with respect to dilemma type and frame between subjects.

FIRST PAGE:

Thank you for participation in this study.

In the following, you find descriptions of situations and subsequent questions.

Please read these descriptions thoroughly and answer the questions according to the instructions.

Please note:

- The data collected is completely anonymous.
- There are no “right” or “wrong” answers.
- Do not talk to each other while answering the questions.
- The person next to you answers different questions.
- **Please, do not turn back pages** after you proceeded to the next question.

Bystander/*negative (positive)*:

The brakes of a moving train have failed. Unfortunately, there are five persons in its way on the track. If no action is taken, these five persons *will die with a probability of **[…]%** in (have a **[…]%** chance to survive)* this incident. One person is located near a lever beside the track, which he can use to operate a switch. This would result in the train being diverted onto a sidetrack before it reaches the five persons. However, there is a single person on this track. If the lever is pulled and the train diverted, this person *will die with a probability of **[…]%** in (has a **[…]%** chance to survive)* this incident.

How morally permissible is it to pull the lever and divert the train?

Please answer by means of the following scale,

with **0** representing **complete impermissibility**

and **5** representing **complete permissibility**

You can use the values in between to graduate your judgement.

completely impermissible					completely permissible
0	1	2	3	4	5

Study 1	**Study 2**	**Study 3**
1.1 100% (0%); 100% (0%)	**2.1** 100% (0%); 100% (0%)	**3.1** 50% (50%); 100% (0%)
1.2 80% (20%); 100% (0%)	**2.2** 99% (1%); 99% (1%)	**3.2** 49% (51%); 98% (2%)
1.3 60% (40%); 100% (0%)	**2.3** 80% (20%); 80% (20%)	**3.3** 40% (60%); 80% (20%)
1.4 40% (60%); 100% (0%)	**2.4** 60% (40%); 60% (40%)	**3.4** 30% (70%); 60% (40%)
1.5 20% (80%); 100% (0%)	**2.5** 40% (60%); 40% (60%)	**3.5** 20% (80%); 40% (60%)
1.6 10% (90%); 100% (0%)	**2.6** 20% (80%); 20% (80%)	**3.6** 10% (90%); 20% (80%)
	2.7 1% (99%); 1% (99%)	**3.7** 1% (99%); 2% (98%)

Footbridge/*negative (positive)*:

The brakes of a moving train have failed. Unfortunately, there are five persons in its way on the track. If no action is taken, these five persons *will die with a probability of **[…]%** in (have a **[…]%** chance to survive)* this incident. On a footbridge above the track, there is a person next to another, very big person. The only way to stop the train before it reaches the five persons is to shove the big person onto the track. If the train is stopped by shoving the big person on the track, this person *will die with a probability of **[…]%** in (has a **[…]%** chance to survive)* this incident.

How morally permissible is it to shove the person onto the track to stop the train?

Please answer by means of the following scale,

with **0** representing **complete impermissibility**

and **5** representing **complete permissibility**

You can use the values in between to graduate your judgement.

completely impermissible					completely permissible
0	1	2	3	4	5

Study 1	**Study 2**	**Study 3**
1.1 100% (0%); 100% (0%)	**2.1** 100% (0%); 100% (0%)	**3.1** 50% (50%); 100% (0%)
1.2 80% (20%); 100% (0%)	**2.2** 99% (1%); 99% (1%)	**3.2** 49% (51%); 98% (2%)
1.3 60% (40%); 100% (0%)	**2.3** 80% (20%); 80% (20%)	**3.3** 40% (60%); 80% (20%)
1.4 40% (60%); 100% (0%)	**2.4** 60% (40%); 60% (40%)	**3.4** 30% (70%); 60% (40%)
1.5 20% (80%); 100% (0%)	**2.5** 40% (60%); 40% (60%)	**3.5** 20% (80%); 40% (60%)
1.6 10% (90%); 100% (0%)	**2.6** 20% (80%); 20% (80%)	**3.6** 10% (90%); 20% (80%)
	2.7 1% (99%); 1% (99%)	**3.7** 1% (99%); 2% (98%)

LAST PAGE:

How do you judge yourself:

Are you generally risk tolerant or do you try to avoid risks?

Please answer by means of the following scale,

with **0** representing **no risk tolerance**

and **10** representing **high risk tolerance**

You can use the values in between to graduate your judgement.

no risk tolerance										high risk tolerance
0	1	2	3	4	5	6	7	8	9	10

Finally, please provide some additional information:

Gender:

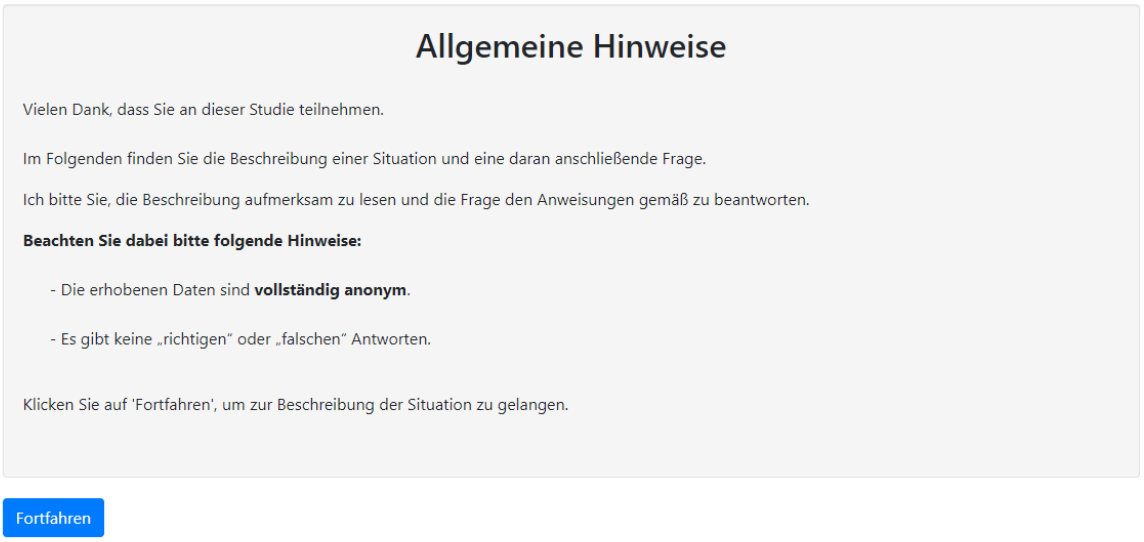
Age:

Course of studies:

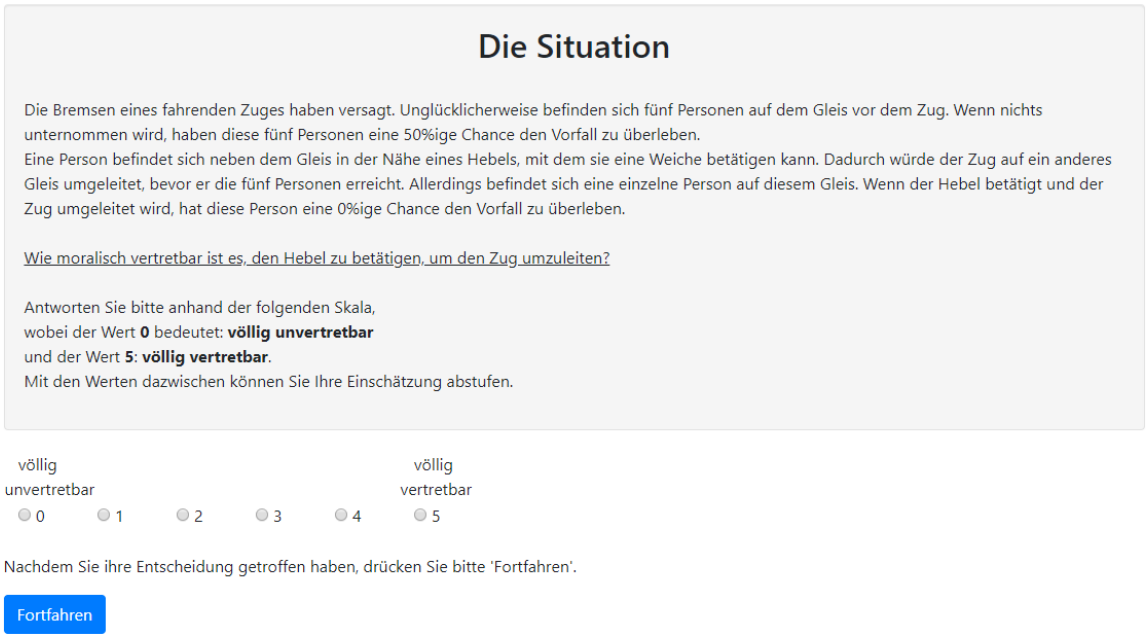
Experimental instructions study 4

Subjects saw a total of three screens. The first screen and the last screen were the same for all subjects. The screen in between contained one dilemma each. The translations provided before apply accordingly.

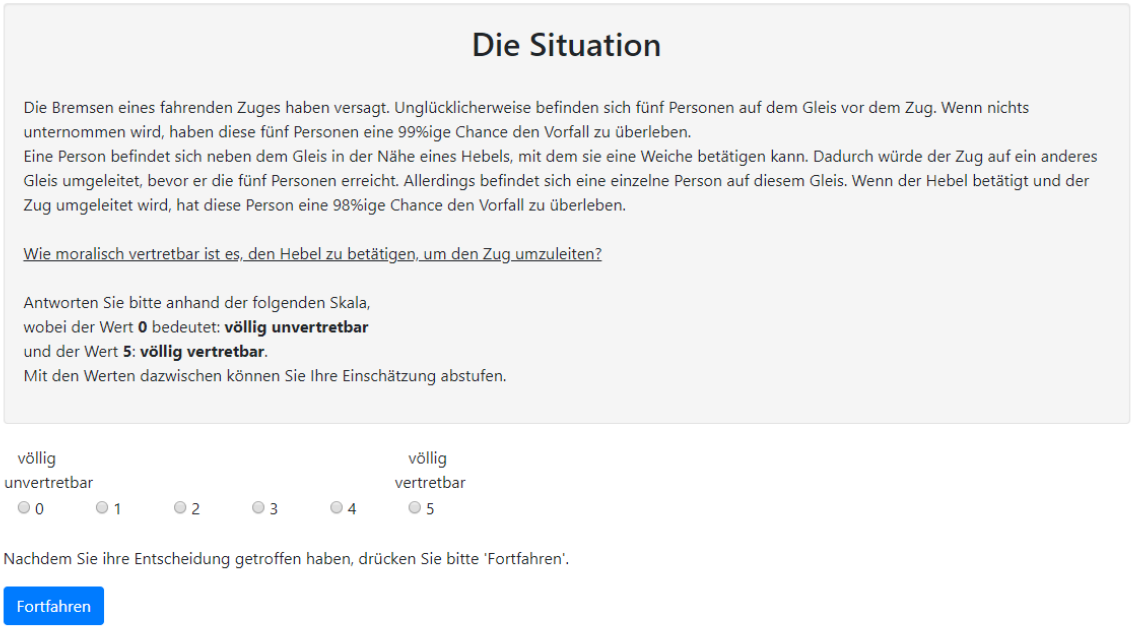
FIRST SCREEN:



3.1 *bystander/positive*



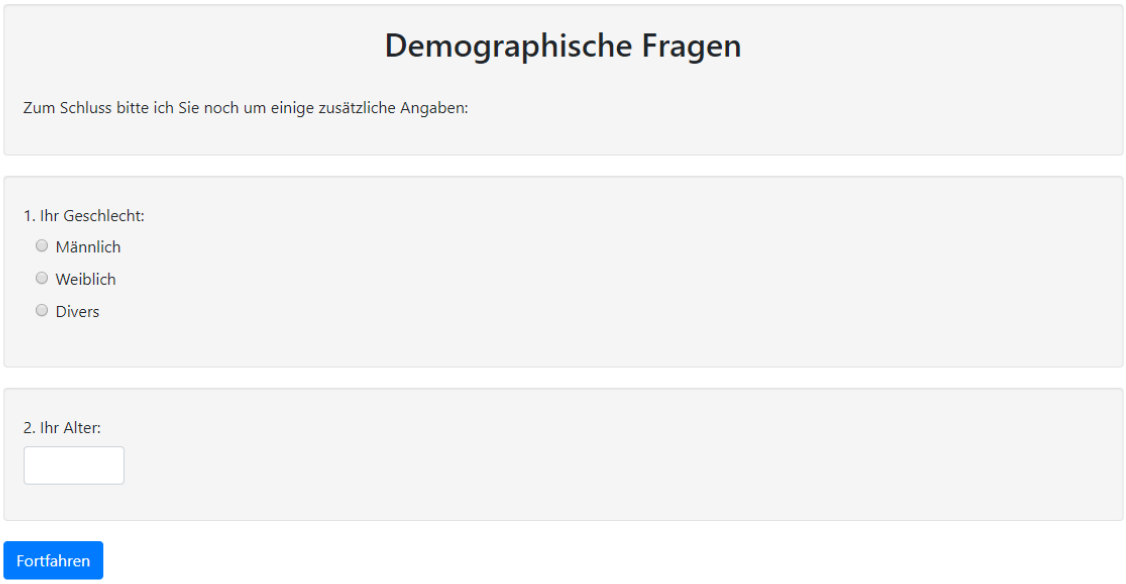
3.7 *bystander/positive*:



3.1 *footbridge/positive*

3.7 *footbridge/positive*:

LAST SCREEN:



4 Framing and Gender Effects in the Sender-Receiver Game

Abstract

This research employs four variants of the standard sender-receiver game by Gneezy (2005), with outcome valence being varied systematically. Depending on the frame of the game, a deceptive message, if acted upon, resulted in a higher gain for the sender and a lower gain for the receiver, a lower loss for the sender and a lower gain for the receiver, a higher gain for the sender and a higher loss for the receiver or a lower loss for the sender and a higher loss for the receiver. Results show that framing has no effect on senders' decisions to lie on aggregate. Analyses with respect to gender point towards female and male subjects being influenced differently by the framing manipulation. Women show a higher propensity to lie to avoid a higher loss and behave less deceptively if doing so increases the receiver's loss, with this pattern being reversed for men.

4.1 Introduction

Many economic activities exhibit a disparate allocation of relevant information among the involved parties. In fact, one might even argue that a situation where all information, including all possible payoffs resulting from different activities, is available to all actors to the exact same extent is an outright exception rather than the rule. Illustrative examples in this regard readily come to mind. One might think of an investment advisor proposing a certain pension insurance or stock option to a client, a doctor suggesting a certain treatment for a patient, a car salesman advertising used cars and basically all situations where the results of the available options are not immediately obvious to the person making the decision who usually depends on expert knowledge provided by the other party. In light of this premise, the possibility of behaving dishonestly by withholding or misreporting private information is of vast importance with regard to economic outcomes. Previous empirical research has shown that individuals do indeed exploit informational advantages by use of deception, lies, cheating and generally unethical behavior in order to increase their own material benefit to a substantial degree. Yet there is also strong and convincing evidence that a preference for truth telling and honesty significantly shapes many individuals' behavior, so that the prevalence of ethically questionable conduct turns out far below what standard economic theory, characterizing economic agents as solely self-interested, would predict (for reviews see Rosenbaum et al., 2014 and Abeler et al., 2019).

The motivation for the research at hand is the idea that outcome valence might make a notable difference for the decision to engage in deceptive behavior. More precisely, I hypothesize that on the one hand, individuals have a stronger incentive to behave unethically, if the goal is reducing one's own losses rather than increasing gains. This conjecture roots in the literature on loss aversion, i.e., the notion that losses have a stronger impact on individual utility than gains of equal size (Kahneman and Tversky, 1979, 1984; Tversky and Kahneman, 1981, 1991, 1992). Although its universal validity has been questioned recently (Gal and Rucker, 2018, Walasek et al., 2018), the concept of loss aversion is among the most widely accepted and utilized in behavioral economics. On the other hand, I assume that individuals show an increased reluctance to deceive their counterpart, if doing so results in higher losses for the other party rather than in foregone gains. This notion is informed by evidence of the existence of a "do-no-harm" principle (Baron, 1996), that is to be found in the literature on bargaining (Tornblom and Jonsson, 1985, Messick and Schell, 1992, van Beest et al., 2005, Leliveld et al., 2009).

There is some previous evidence on the influence of framing outcomes as losses as compared to gains on unethical behavior.

Cameron and Miller (2009) show that subjects cheat more when reporting the number of correctly solved anagrams in a loss frame compared to a gain frame. Kern and Chugh (2009) give account of subjects being more inclined to gather "insider information" in a loss frame and of lying in negotiations if a loss frame is induced. Grolleau et al. (2016) report that individuals overstate their performance in

a real-effort matrix task more strongly in a loss frame compared to a gain frame. Schindler and Pfattheicher (2017) show that individuals cheat more when reporting outcomes of coin tosses and dice rolls in a loss frame compared to a gain frame. Most closely related to this research are the studies by Reinders Folmer and De Cremer (2012) and Childs (2012). Reinders Folmer and De Cremer report that individuals are more inclined to lie to their counterpart about the actual endowment to be allocated in an ultimatum game in a loss frame rather than a gain frame and can show that behavior of subjects characterized as proselfs by an SVO task (van Lange et al., 1997; van Lange 1999) induces this difference. Childs finds no framing effects in a loss frame in a high stake sender-receiver game compared to a gain frame.

My study contributes to the literature by rigorously disentangling the two presumably opposing effects of outcome valence on deceptive behavior by employing four treatments with different outcome framing in the standard sender-receiver game by Gneezy (2005). This is a simple cheap-talk game with two players, where one player knows the payoffs of two possible options that the second player can choose from. The only information the second player gets is a message from her counterpart telling her which option results in a higher payoff to herself.

In the gain frame, which can be considered the baseline treatment, a lie[1] (if acted upon by the receiver) results in a higher gain for the sender and in foregone gains for the receiver. In the first of two mixed frames, which is employed to detect the possible existence of a loss-aversion effect, a lie results in a smaller loss for the sender and foregone gains for the receiver. In the second mixed frame, aimed at testing for a do-no-harm effect, this is reversed and a lie results in a higher gain for the sender and a higher loss for the receiver. In the loss frame the two aforementioned effects are supposed to clash and a lie yields a smaller loss for the sender and a higher loss for the receiver.
Appropriate adjustments of initial endowments ensure that all treatments are payoff equivalent.

This research also contributes to the question, whether there are substantial gender effects in such a sender-receiver game environment. These effects may emerge on two dimensions in this regard, namely there might be differences in the propensity to lie and in the propensity to exhibit trust, i.e., to follow the sender's recommendation. While Dreber and Johannesson (2008) find that males lie significantly more than females, Aoki et al. (2010), Erat and Gneezy (2012), Childs (2012) and Gylfason et al. (2013) fail to establish a significant relationship between gender and the propensity to lie. However, in a meta-analysis that also includes sender-receiver games that extend Gneezy's original design, Capraro (2018) finds that males are significantly more likely to lie than females, yet he explicitly acknowledges that this effect is fairly small and only shows statistical significance when using very large

[1] The notion of a lie in this paper is consistently the notion of a selfish black lie as defined by Erat and Gneezy (2009), i.e., one that is used to increase the liar's payoff at the expense of another person.

samples. Dreber and Johannesson (2008), Aoki et al. (2010), Childs (2012) and Gylfason et al. (2013) report that they do not find gender differences with respect to trust exhibited.

The remainder of this paper is organized as follows. Section 2 depicts the experimental design and procedure. Section 3 states the hypotheses. Section 4 shows the results and section 5 provides a discussion and concludes.

4.2 Design and procedure

To elicit the effects of outcome valence on the decision to behave deceptively, I employed four treatments in Gneezy's standard sender-receiver game (2005) in which outcomes were framed differently. I followed Gneezy's original protocol closely and conducted this research as pen and paper experiments in the classroom. Participants were students enrolled in an introductory economics class at Kiel University, Germany. They were approached after the tutorials accompanying the lecture and asked to take part in a short experiment in decision-making. All subjects got a 2€ participation fee in addition to their earnings from the experiment. They received their total payments one week after their actual decisions. I first collected 128 choices of senders, who were randomly assigned to one of the four treatments, with 32 observations in each of these. Senders were informed that they would be anonymously matched with another person (the receiver) in another room, who has to decide between two options (A and B), that result in different payments for both parties. Senders were fully informed about the options' consequences, while it was common knowledge that the receiver would not have this information but merely get a message from the sender, stating which option yields a higher final payoff for the receiver. Specifically, senders could choose between message 1: "*Option A will earn you more money than option B."* and message 2: "*Option B will earn you more money than Option A*". The chosen messages were forwarded to 128 receivers, who then made the decisions that determined final payments (refer to the Appendix for the exact experimental instructions).

The first treatment (referred to as GAIN henceforth) is basically a replication of the standard setting, with both options resulting in additional gains for the two parties. Additional payments were gains of 2€ and 4€ either in favor of the Sender (option A) or the receiver (option B).

In the first of two mixed frame treatments (MIXED 1), endowments were introduced[2] such that the senders were initially endowed with 6€, while the receivers did not receive any endowment. Now both options resulted in losses for the sender (2€ with option A, 4€ with option B), while the receiver was paid additional 2€ with option A and gained 4€ with option B.

For the second mixed frame (MIXED 2), endowments were reversed, such that the senders had no endowment, while the receivers had 6€. Option A (B) now resulted in a gain of 4€ (2€) for the sender and a loss of 4€ (2€) for the receiver.

[2] Initial endowments were common knowledge and not senders' private information.

In the loss treatment (LOSS), both players were endowed with 6€ and both options resulted in losses of 2€ and 4€, either in favor of the sender (option A) or the receiver (option B).

Table 1 summarizes this information and emphasizes that all treatments are completely payoff equivalent and only differ with respect to outcome framing.

I reversed the payments associated with options A and B for half of the subjects in each treatment, such that sending message 1 implied behaving truthfully instead of lying. This was obviously meant to counteract possible order effects or potential biases brought about by innate proclivities to prefer one message over the other.

Table 1: Payoff structures by treatment

Treatment	Endowments	Option	Additional payments	Final payoffs
GAIN	*Sender*: 0€ *Receiver*: 0€	A	*S*: 4€ gain, *R*: 2€ gain	*S*: 4€, *R*: 2€
		B	*S*: 2€ gain, *R*: 4€ gain	*S*: 2€, *R*: 4€
MIXED 1	*Sender*: 6€ *Receiver*: 0€	A	*S*: 2€ loss, *R*: 2€ gain	*S*: 4€, *R*: 2€
		B	*S*: 4€ loss, *R*: 4€ gain	*S*: 2€, *R*: 4€
MIXED 2	*Sender*: 0€ *Receiver*: 6€	A	*S*: 4€ gain, *R*: 4€ loss	*S*: 4€, *R*: 2€
		B	*S*: 2€ gain, *R*: 2€ loss	*S*: 2€, *R*: 4€
LOSS	*Sender*: 6€ *Receiver*: 6€	A	*S*: 2€ loss, *R*: 4€ loss	*S*: 4€, *R*: 2€
		B	*S*: 4€ loss, *R*: 2€ loss	*S*: 2€, *R*: 4€

4.3 Hypotheses

The hypotheses about sender behavior follow straightforwardly from the conjectures derived previously. If subjects are loss averse, they will be more content to engage in deceptive behavior in order to reduce an own loss rather than to increase a gain, thus the first hypothesis reads:

H1a: *The fraction of senders conveying a deceptive message is larger in MIXED 1 than in GAIN.*

Similarly, if increasing the receiver's loss by use of deception is perceived less ethically acceptable than reducing her gain by lying, senders are less ready to behave deceptively in the first case, such that the second hypothesis is:

H1b: *The fraction of senders conveying a deceptive message is smaller in MIXED 2 than in GAIN.*

The third hypothesis rests on the assumption that the two assumed effects in the mixed treatments act contrary to each other if put to work simultaneously in LOSS and therefore reads:

H1c: *The fraction of senders conveying a deceptive message in LOSS is between those in the two mixed treatments.*

As receivers do not have any information on the actual additional payoffs associated with the two options, I hypothesize that their choices are not influenced by treatment differences.

H2: *The fraction of those who exhibit trust does not differ between treatments.*

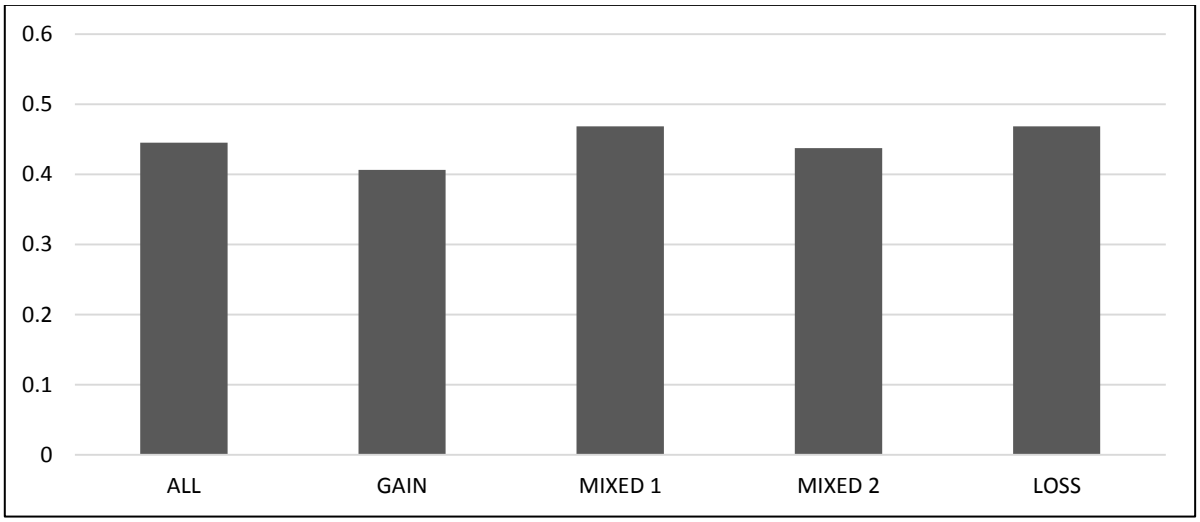
With respect to gender, one additional hypothesis can be derived from insights with respect to loss aversion for risky choices. Previous empirical work in this field has shown that females exhibit a greater degree of loss aversion than males (Schmidt and Traub, 2002, Rau, 2014, Hassan et al., 2014, Arora and Kumari, 2015), such that I derive the following hypothesis:

H3: *Females' decisions to lie are more likely to be encouraged by the treatment manipulation in MIXED 1 than males'.*

4.4 Results

Figure 1 illustrates the fraction of senders who decided to send a deceptive message in each of the four treatments and also includes a joint representation (ALL).

Figure 1: Fraction of lies by treatment



In GAIN (n=32, 50% female, mean age 20.8 years) 13 senders (40.6%) lied. In MIXED 1 (n=32, 50% female, mean age 21.1 years) as well as in LOSS (n=32, 50% female, mean age 20.8 years), 15 senders (46.9%) send a deceptive message. In MIXED 2 (n=32, 62.5% female, mean age 20.7 years), 14 senders (43.8%) send false information. This adds up to a total number of 57 senders (44.5%) lying. These fractions are well in line with previous observations that have been reported (e.g., 36% in Gneezy, 2005, 47% in Dreber and Johannesson, 2008, 44% in Sutter, 2009 and 44% in Gylfason et al., 2013). It is also readily observable that the fractions of senders behaving untruthfully do not differ substantially

between treatments and pairwise tests of differences of proportions support this impression. Table 2 provides the respective p-values of all these two-sided tests.

Table 2: Fraction lies, pairwise comparisons between treatments

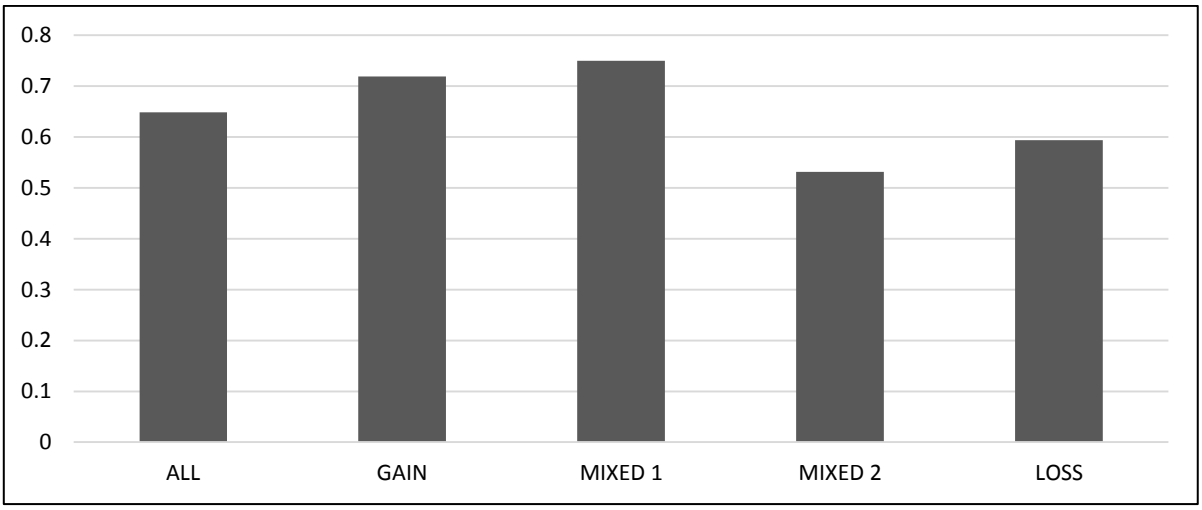
Fraction lies	MIXED 1 (0.469, 15/32)	MIXED 2 (0.438, 14/32)	LOSS (0.469, 15/32)
GAIN (0.406, 13/32)	p=0.614	p=0.800	p=0.614
MIXED 1 (0.469, 15/32)	-	p=0.802	p=1.000
MIXED 2 (0.438, 14/32)	-	-	p=0.802

*Note: two-sided tests of differences in proportions; *=p<0.1, **=p<0.05, ***=p<0.01*

Hence, I do not find any support for hypotheses H1a-c on aggregate.

Now turning to the receivers' behavior, figure 2 depicts the fractions of subjects exhibiting trust, i.e., actually following the sender's assertion and choosing the option that allegedly results in the highest payoff for themselves.

Figure 2: Fraction of trust exhibited by treatment



In GAIN (n=32, 43.8% female, mean age 21.9 year), 23 subjects (71.9%) exhibited trust. In MIXED 1 (n=32, 34.4% female, mean age 21.1 years) 24 receivers (75%) followed the sender's assertion. In MIXED 2 (n=32, 50% female, mean age 21.2 years) 17 subjects (53%) chose the option ostensibly associated with a higher payoff, while 19 (59.4%) do so in LOSS (n=32, 34.4% female, mean age 19.8 years). This makes for a total of 83 receivers (64.8%) exhibiting trust across all treatments. These results support previous findings that a majority of receivers follow the senders' lead (e.g., 78% in Gneezy, 2005, 76.1% in Dreber and Johannesson, 2008, 72.7% in Childs, 2012, 77.7% in Gylfason et al., 2013),

yet I also observe some considerable variation between treatments in this regard. Table 3 shows the p-values of pairwise two-sided tests for differences of proportions between treatments.

Table 3: Fraction trust exhibited, pairwise comparisons between treatments

Fraction trust	MIXED 1 (0.75, 24/32)	MIXED 2 (0.53, 17/32)	LOSS (0.594, 19/32)
GAIN (0.719, 23/32)	p=0.777	p=0.121	p=0.293
MIXED 1 (0.75, 24/32)	-	p=0.068*	p=0.183
MIXED 2 (0.53, 17/32)	-	-	p=0.614

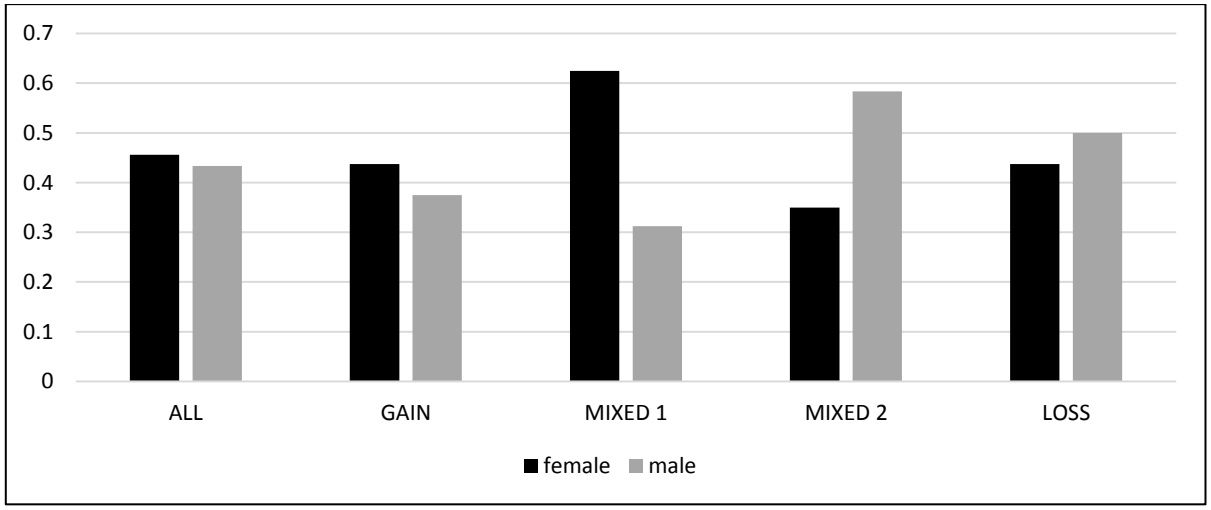
*Note: two-sided tests of differences in proportions; *=p<0.1, **=p<0.05, ***=p<0.01*

While the only significant difference emerges when comparing MIXED 1 and MIXED 2 with two sided-test, additional one sided-tests reveal a tendency that trust exhibited in GAIN is higher than in MIXED 2 (p=0.061), and also higher in MIXED 1 than in LOSS (p=0.092). Employing a one-sided test also increases the level of significance when comparing the two mixed treatments (p=0.034). In total, I regard this as a sufficient basis to reject H2, i.e., the assertion of equivalent fractions.

Gender effects

While I do not observe any considerable effect of framing on the decision to convey a deceptive message on aggregate, analyzing choices of female and male subjects separately diversifies the picture. As to be seen in figure 3, that depicts fractions of lying senders by gender in all treatments, again including a pooled representation, the decision to behave deceptively seems to be influenced by framing for both genders.

Figure 3: Fraction lies by treatment and gender



Female subjects show reactions to different frames as hypothesized in the beginning. Proclivity to lie is highest in MIXED 1, lowest in MIXED 2 and medium in GAIN and LOSS. The exact opposite pattern emerges for male subjects, with the highest fraction of deceptive behavior in MIXED 2 and the lowest in MIXED 1.

As Table 4 shows, none of the differences between treatments by gender shows statistical significance with two-sided tests.

Table 4: Fraction lies, pairwise comparisons between treatments by gender

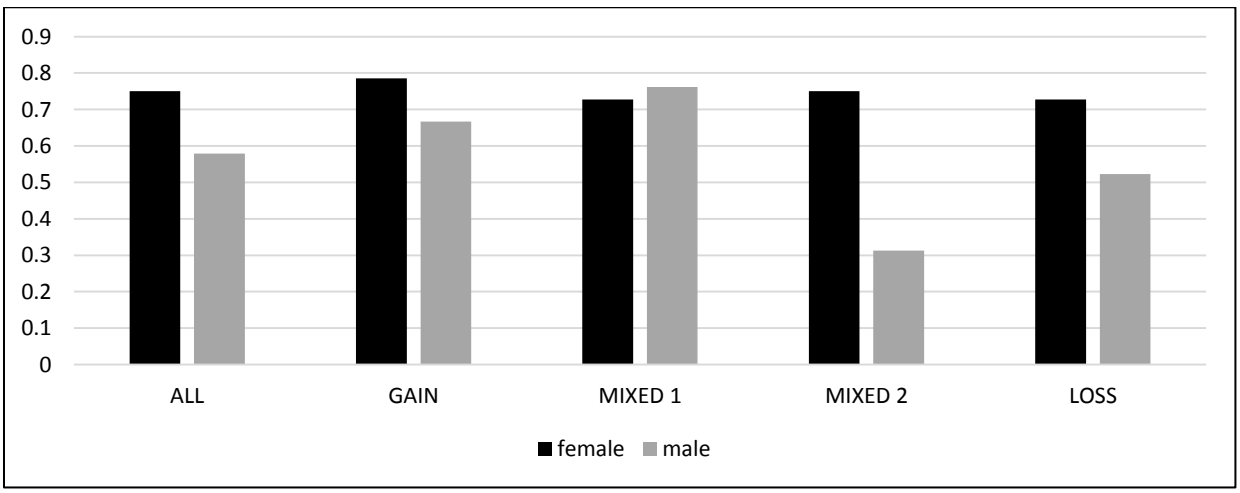
Fraction lies	MIXED 1	MIXED 2	LOSS
GAIN	female: p=0.288	female: p=0.593	female: p=1.000
	male: p=0.710	male: p=0.275	male: p=0.476
MIXED 1	-	female: p=0.101	female: p=0.288
	-	male: p=0.152	male: p=0.280
MIXED 2	-	-	female: p=0.593
	-	-	male: p=0.663

*Note: two-sided tests of differences in proportions; *=p<0.1, **=p<0.05, ***=p<0.01*

However, employing additional one-sided tests renders the differences between the two mixed treatments weakly significant for both groups (female: p=0.0503; male: p=0.0762), which I, with all due care, interpret as my framing manipulation being effective to some degree. More precisely, I regard this as frail evidence that H1a-c hold for females, while the exact opposite of H1a and H1b is true for males. However, the LOSS treatment also yields a result that lies between the two mixed treatments for the latter group, which is in line with H1c.

Figure 4 depicts the fractions of those receivers who followed the senders lead in all treatments by gender.

Figure 4: Fraction trust exhibited by treatment and gender



Apparently, framing does not substantially influence female subjects in this regard. This impression is backed up by the results of pairwise, two-sided tests of differences in proportions that are depicted in Table 5, such that H2 is supported for female subjects. A different picture emerges if analyzing trust exhibited by males in the different treatments. The fraction of trusting males is significantly higher in GAIN and MIXED 1 than in MIXED 2 and additional one-sided tests point at a weakly significant tendency that trust is higher in MIXED 1 than in LOSS (p=0.087) and lower in MIXED 2 compared to LOSS (p=0.099), which leads to rejection of H2 for male subjects.

Table 5: Fraction trust exhibited, pairwise comparisons between treatments by gender

Fraction trust	MIXED 1	MIXED 2	LOSS
GAIN	female: p=0.887	female: p=0.816	female: p=0.732
	male: p=0.684	male: p=0.039**	male: p=0.365
MIXED 1	-	female: p=0.943	female p=0.851
	-	male: p=0.012**	male: p=0.174
MIXED 2	-	-	female: p=0.893
	-	-	male: p=0.198

*Note: two-sided tests of differences in proportions; *=p<0.1, **=p<0.05, ***=p<0.01*

Finally, I report a closer investigation of gender differences by treatments. Table 6 shows the results of the respective comparisons for both, senders and receivers.

Table 6: Differences in proportions between genders by treatment

Treatment	**Fraction**	**Female**	**Male**	**p-value**
All	lies	0.456 (31/68)	0.433 (26/60)	0.798
	trust	0.75 (39/52)	0.579 (44/76)	0.047**
GAIN	lies	0.4375 (7/16)	0.375 (6/16)	0.719
	trust	0.786 (11/14)	0.667 (12/18)	0.458
MIXED 1	lies	0.625 (10/16)	0.3125 (5/16)	0.077*
	trust	0.727 (8/11)	0.762 (16/21)	0.830
MIXED 2	lies	0.35 (7/20)	0.583 (5/12)	0.198
	trust	0.75 (12/16)	0.3125 (5/16)	0.013**
LOSS	lies	0.4375 (7/16)	0.5 (8/16)	0.723
	trust	0.727 (8/11)	0.524 (11/21)	0.266

*Note: two-sided tests of differences in proportions; *=p<0.1, **=p<0.05, ***=p<0.01*

While the overall fraction of liars does not differ significantly, I observe a considerable difference in the propensity to engage in deceptive behavior in MIXED 1. Additional one-sided tests also reveal a weakly significant difference in MIXED 2 (p=0.099) and render the difference in MIXED 1 significant at the 5%-level (p=0.038), which is in line with H3.

Overall trust differs significantly, with this difference being mainly driven by the extraordinarily low fraction of trusting males in MIXED 2, which of course also results in a significant gender difference in behavior in this particular treatment.

4.5 Discussion

The main conclusion of this research is clearly that framing as employed does not yield the hypothesized effects on deceptive behavior on aggregate. A general robustness of preferences over lying, i.e., the assumption that a certain fraction of individuals always lies if it maximizes their own material payoff, cannot account for this. Gneezy (2005) has already shown in his seminal contribution that the fraction of unethical behavior is influenced by both, efficiency concerns and the amount to be gained from the lie, and others have identified additional determinants shaping behavior, e.g., the size of the lie (Lundquist et al., 2009), the personalization of messages (Cappelen et al., 2013) or the time of day (Kouchaki and Smith, 2014).

The fact that trust exhibited by receivers varies between treatments is somewhat intriguing and may be understood by resorting to the existence of some kind of experimenter demand effect (Zizzo, 2010). Assuming that receivers anticipate possible payoffs and senders' behavior might explain why trust exhibited differs, most pronounced for males in MIXED 2, where the fraction of trust is lowest and the fraction of lies highest. On the other hand, I do observe fractions of trust exhibited that are above 50% for all treatments, such that a sender, who correctly anticipates this, would not have an incentive to engage in "sophisticated deception" (Sutter, 2009), i.e., to tell the truth if he anticipates that the receiver disbeliefs. These beliefs, however, have not been collected, such that the absence of such behavior is merely a conjecture rather than an actual result.

It is also apparent in general that gender seems to play a vital role with regard to the effects of my framing manipulations. While I do not find gender differences in the baseline GAIN treatment, which is in line with most of the previous literature on this matter, both mixed frames appear to provoke diametrically opposed reactions between genders, with behavior of females supporting my hypotheses. To be sure, words of caution are in order when drawing inferences from the analyses with respect to gender, as the sample size is fairly small and statistical power thus naturally limited. However, these initial results may pave the way for future research in this field.

Another limitation of this study is the fact that all possible payoffs for one person in the different treatments do not differ with respect to outcome valence, i.e., the choice is neither between generating a gain or not, nor between bearing or inflicting a loss or not. The only difference between the options is the size of the gain or loss. It appears worthwhile to test whether it makes a difference when a loss or gain can be avoided altogether.

References

Abeler, J., Nosenzo, D. & Raymond, C. (2019). Preferences for Truth-Telling. Econometrica, 87(4), 1115-1153.

Aoki, K., Akai, K. & Onoshiro, K. (2010). Deception and confession: Experimental evidence from a deception game in Japan. ISER Discussion Paper, No. 786.

Arora, M. & Kumari, S. (2015). Risk Taking in Financial Decisions as a Function of Age, Gender: Mediating Role of Loss Aversion and Regret. *International Journal of Applied* Psychology, 5(4), 83-89.

Baron, J. (1996). Do no harm. In D. M. Messick & A. E. Tenbrusel (Eds.), *Codes of conduct: Behavioral research into business ethics*, pp. 197-213, New York: Russell Sage Foundation.

Cameron, J. S. & Miller, D. T. (2009). Ethical standards in gain versus loss frames. In. D. de Cremer (Ed.), *Psychological perspectives on ethical behavior and decision making*, pp. 91-106, Charlotte, NC: Information Age Publishing.

Cappelen, A. W. Sorensen, E. O., Tungodden, B. (2013). When do we lie? *Journal of Economic Behavior and Organization*, 93, 258-265.

Capraro, V. (2018). Gender Differences in Lying in Sender-Receiver Games: A Meta-Analysis. *Judgement and Decision Making*, 13(4), 345-355.

Childs, J. (2012). Gender differences in lying. *Economics Letters*, 114, 147-149.

Dreber, A. & Johannesson, M. (2008). Gender differences in deception. *Economics Letters*, 99, 197-199.

Erat, S. & Gneezy, U. (2012). White Lies. *Management Science*, 58, 723-733.

Gal, D. & Rucker, D. D. (2018). The Loss of Loss Aversion: Will It Loom Larger Than Its Gain? *Journal of Consumer Psychology*, 28(3), 497-516.

Gneezy, U. (2005). Deception: The Role of Consequences. *The American Economic Review*, 95(1), 384-394.

Grolleau, G., Kocher, M. G. & Sutan, A. (2016). Cheating and Loss Aversion. Do People Cheat More to Avoid a Loss? *Management Science*, 62(12), 3428-3438.

Gylfason, H. F., Arnardottir, A. A. & Kristinsson, K. (2013). More on gender differences in lying. *Economics Letters*, 119, 94-96.

Hassan, R. H., Khalid, W. & Habib, A. (2014). Overconfidence and Loss Aversion in Investment Decisions: A Study of the Impact of Gender and Age in Pakistani Perspective. *Research Journal of Finance and Accounting*, 5(11), 148-157.

Kahneman, D. & Tversky, A. (1979). Prospect Theory: An analysis of decision under risk. *Econometrica*, 47, 263-292.

Kahneman, D. & Tversky, A. (1984). Choices, values, and frames. *The American Psychologist*, 39, 341-350.

Kern, M. C. & Chugh, D. (2009). Bounded Ethicality: The Perils of Loss Framing. *Psychological Science*, 20(3), 378-384.

Kouchaki, M. & Smith, I. H. (2014). The morning morality effect: The influence of time of day on unethical behavior. *Psychological Science*, 25, 95-102.

Leliveld, M. C., Van Beest, I., Van Dijk, E. & Tenbrusel, A. E. (2009). Understanding the influence of outcome valence in bargaining: A study on fairness accessibility, norms, and behavior. *Journal of Experimental Social Psychology*, 45, 505-514.

Lundquist, T., Ellingsen, T., Gribbe, E. & Johannesson, M. (2009). The aversion to lying. *Journal of Economic Behavior and Organization*, 70, 81-92.

Messick, D. M., & Schell, T. (1992). Evidence for an equality heuristic in social decision making. *Acta Psychologica*, 80, 311-323.

Rau, H. A. (2014). The disposition effect and loss aversion: Do gender differences matter? *Economics Letters*, 123, 33-36.

Reinders Folmer, C. P. & De Cremer, D. (2012). Bad for Me or Bad for Us? Interpersonal Orientations and the Impact of Losses on Unethical Behavior. *Personality and Social Psychology Bulletin*, 38(6), 760-771.

Rosenbaum, S. M., Billinger, S. & Stieglitz, N. (2014). Let's be honest: A review of experimental evidence of honesty and truth-telling. *Journal of Economic Psychology*, 45, 181-196.

Schindler, S. & Pfattheicher, S. (2017). The frame of the game: Loss-framing increases dishonest behavior. *Journal of Experimental Social Psychology*, 69, 172-177.

Schmidt, U. & Traub, S. (2002). An Experimental Test of Loss Aversion. *The Journal of Risk and Uncertainty*, 25(3), 233-249.

Sutter, M. (2009). Deception through telling the truth?! Experimental evidence from individuals and teams. *Economic Journal*, 119, 47-60.

Tornblom, K. Y. & Jonsson, D. R. (1985). Subrules of the Equality and Contribution Principles: Their Perceived Fairness in Distribution and Retribution. *Social Psychology Quarterly*, 48(3), 249-261.

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

Tversky, A. & Kahneman, D. (1991). Loss aversion in riskless choice: A reference-dependent model. *Quarterly Journal of Ecoomics*, 106, 1039-1061.

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

Van Beest, I., Van Dijk, E., De Dreu, C. K. W. & Wilke, H. A. M. (2005). Do-no-harm in coalition formation: Why losses inhibit exclusion and promote fairness cognitions. *Journal of Experimental Social Psychology*, 41, 609-617.

Van Lange, P. A. M. (1999). The Pursuit of Joint Outcomes and Equality in Outcomes: An Integrative Model of Social Value Orientation. *Journal of Personality and Social Psychology*, 77(2), 337-349.

Van Lange, P. A. M., De Bruin, E. M. N., Otten, W. & Joireman, J. M. (1997). Development of Prosocial, Individualistic and Competitive Orientations: Theory and Preliminary Evidence. *Journal of Personality and Social Psychology*, 73(4), 733-746.

Walasek, L., Mullett, T. L. & Stewart, N. (2018) A Meta-Analysis of Loss Aversion in Risky Contexts *Unpublished Manuscript.* Available at SSRN: https://ssrn.com/abstract=3189088 or http://dx.doi.org/10.2139/ssrn.3189088

Zizzo, D. J. (2010). Experimenter demand effects in economic experiments. *Experimental Economics*, 13(1), 75-98.

Appendix

Experimental instructions, sender (translated from German; parts that differ between treatments are depicted in italics):

Thank you for participation in this short experiment in decision-making.

As all other participants, you receive 2€ for participation. You can acquire additional money depending on another person's decision, who is in another room.

Neither of you will ever get information about the other's identity.

Your total payment will be handed to you in cash after next week's tutorial.

GAIN:

Two options are available that result in additional payoffs for you and your counterpart.

Option A: additional 4€ for you and additional 2€ for your counterpart

Option B: additional 2€ for you and additional 4€ for your counterpart

MIXED 1:

You are endowed with 6€, while your counterpart does not have any endowment.

Two options are available that result in additional payoffs for you and your counterpart.

Option A: a loss of 2€ for you and additional 2€ for your counterpart

Option B: a loss of 4€ for you and additional 4€ for your counterpart

MIXED 2:

You don't get any endowment, while your counterpart is endowed with 6€.

Two options are available that result in additional payoffs for you and your counterpart.

Option A: additional 4€ for you and a loss of 4€ for your counterpart

Option B: additional 2€ for you and a loss of 2€ for your counterpart

LOSS

Both, you and your counterpart, are endowed with 6€.

Two options are available that result in additional payoffs for you and your counterpart.

Option A: a loss of 2€ for you and a loss of 4€ for your counterpart

Option B: a loss of 4€ for you and a loss of 2€ for your counterpart

The other person will choose which option is realized. The only information this person has, is a message that you convey. This means that your counterpart does not know the payoffs associated with the two options.

Please decide now which of the following messages you want to send to your counterpart:

Message 1: "Option A will earn you more money than option B."

Message 2: "Option B will earn you more money than option A."

We will show your message to your counterpart. We will then ask your counterpart to choose between option A and option B. This decision will determine the actual payoffs. Your counterpart will never know what sums were offered in the option not chosen. This means that your counterpart will not know whether your message was true or false. Additionally, your counterpart will only know his or her payoff but not yours.

I send the following message (please check the box):

☐ Message 1 ☐ Message 2

Finally, please provide some general information.

Gender:_____ Age:_____ Course of studies:_____

Experimental instructions, receiver (translated from German; parts that differ between treatments are depicted in italics):

Thank you for participation in this short experiment in decision-making.

As all other participants, you receive 2€ for participation. You can acquire additional money depending on your decision.

Your total payment will be handed to you in cash after next week's tutorial.

There are possible additional payoffs for you and another person, who is in another room.

Neither of you will ever get information about the other's identity.

MIXED 1:

You don't get any endowment, while your counterpart is endowed with 6€.

MIXED 2:

You are endowed with 6€, while your counterpart does not have any endowment.

LOSS:

Both, you and your counterpart, are endowed with 6€.

Two options are available that result in additional payoffs for you and your counterpart.

We showed the actual payoffs that result from these options to your counterpart. You do not see these payoffs yourself. The only information you get is a message sent by your counterpart.

Two possible messages could be sent:

Message 1: "Option A will earn you more money than option B."

Message 2: "Option B will earn you more money than option A."

Your counterpart decided to send you message _____.

Please decide now which option you would like to choose. Your decision determines both, your payoff and your counterpart's payoff. You will never know what sums were offered in the option not chosen, and thus not be informed whether the message was true or false. Additionally, you will only know your own payoff and not your counterpart's.

I choose the following option (please check the box):

☐ Option A ☐ Option B

Finally, please provide some general information.

Gender:_____ Age:_____ Course of studies:_____

5 Concluding remarks

By looking at the results of all the studies reported in the dissertation at hand, it becomes obvious that actual individual decision making oftentimes substantially deviates from what one might expect to happen a priori. While some of these empirical observations can reasonably be deemed expectable, e.g., the impact of windfall gains in the first study, the fact that intervention in the footbridge dilemma is judged less permissible than in the bystander case and generally lower if the attraction of intervention is reduced by changing the ratio of expected outcomes, and the insight that a considerable fraction acts honestly in study three even if it counteracts their material self-interest, some other findings may come as a surprise. Among those are the results that prior outcomes do not yield a predictable change of risk attitude in study one, the general lack of framing effects and the manifestation of a common ratio effect in only one of the dilemmas in study two, and the overall lack of framing effects in study three along with the different responses of male and female subjects.

In my view, this calls for two things. First, it is apparently crucial to further develop economics as an empirical and experimental discipline, in order to constantly and tenaciously question the predictions derived from theoretical considerations and to check whether they are confirmed by actual human decision-making. Secondly, one needs to be overly aware that individual decision-making is such a complex phenomenon in its own right that any attempt to model all determinants of behavior with a feasibly scarce set of assumptions is doomed to failure. To be sure, there is a flip side to that coin, namely that it is the very complexity of human behavior that necessitates model-theoretic simplifications. After all, the fact that models simplify is a feature, not a bug. But it is of vast importance to know their boundaries and to keep in mind that and when and at best even why false predictions may be derived. This in turn also means that it would reveal a fundamentally flawed notion of empirical research if one states that experiments “did not work out” in case the collected data is not in line with what has been expected previously. Human behavior cannot be right or wrong in a narrow sense, at least not as investigated in the studies at hand, but rather constitutes a reality that needs to be acknowledged.

Erklärung zum selbständigen Verfassen der Arbeit

Ich erkläre hiermit, dass ich meine Doktorarbeit „Essays on Behavioral Economics: Empirical Studies on Risk, Morality and Framing“ selbständig und ohne fremde Hilfe angefertigt habe und dass ich als Koautor maßgeblich zu dem weiteren Fachartikel beigetragen habe. Alle von anderen Autoren wörtlich übernommenen Stellen, wie auch die sich an die Gedanken anderer Autoren eng anlehnenden Ausführungen der aufgeführten Beiträge wurden besonders gekennzeichnet und die Quellen nach den mir angegebenen Richtlinien zitiert.

Datum Unterschrift