UNIVERSITY OF CALIFORNIA, SAN DIEGO

Emotional Dynamics in Social Behavior and Choice Under Uncertainty

A dissertation submitted in partial satisfaction of the
requirements for the degree
Doctor of Philosophy

in

Economics

by

Alex Oleg Imas

Committee in charge:

Professor Uri Gneezy, Chair
Professor James Andreoni
Professor Ayelet Gneezy
Professor Yuval Rottenstreich
Professor Joel Sobel

2014

The dissertation of Alex Oleg Imas is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

____

____

____

____

____

Chair

University of California, San Diego

2014

DEDICATION

I dedicate this dissertation to my family, who have always been my rock in the often uncertain journey I have followed to this milestone. From fostering my curiosity as a child, to standing behind the choices I made as an adult, thank you for your unconditional support.

TABLE OF CONTENTS

LIST OF FIGURES

ACKNOWLEDGEMENTS

I would like to thank the professors, peers, friends and family who have provided invaluable support and feedback throughout my graduate career, including:

My advisor, Uri Gneezy, who has taught me to seek the important questions and given me the skills to answer them. Since the first quarter of graduate school, he provided me with unwavering support that sparked my career in academia.

Ayelet Gneezy, who fostered my dedication and enthusiasm for research; these vital skills propelled me through the ups and downs of graduate school. In our collaborations, her bottomless energy has been an invaluable inspiration.

Jim Andreoni, who taught me the value of patience in the development and execution of nascent ideas. His advice and constant guidance have been critical in my development as a researcher.

The wonderful faculty at the UCSD Economics Department and Rady, who created an open and exciting environment to nurture my development into a behavioral economist, including Chris Oveis, Yuval Rottenstreich, Sally Sadoff, Pamela Smith and Joel Sobel.

My fellow graduate students, who were always ready to provide helpful feedback, including Matt Knepper, Mike Kuhn, Silvia Saccardo and Larry Schmidt. I would especially like to thank Aislinn Bohren, who provided invaluable support and advice at critical stages of my graduate studies. Our discussions have consistently inspired new projects, improved existing work, and reinvigorated my enthusiasm for economics.

The many other mentors and teachers who have shaped my academic journey, from my high school teachers, Laura Wright and Harry Kyriazes, who ignited my intellectual curiosities, to my undergraduate professors at Northwestern University, including Eric Schulz and Ben Jones, who inspired me to pursue a career in economics.

Kristóf Madarász, who was both an excellent mentor and a wonderful collaborator on the fourth chapter of my dissertation, “Conscience Accounting: Emotion Dynamics in Social Behavior.”

My parents, Shulim and Dora Imas, sister, Polina Imas, and grandmother, Charna Leubelchick who have always stood by me as I pursued new challenges.

I would also like to thank the many individuals and organizations who provided me with financial support, including Uri Gneezy, the Graduate Summer Research Award and the UCSD Institute for Applied Economics. I would especially like to thank the National Science Foundation, whose Graduate Research Fellowship provided funding for the majority of my tenure at UCSD.

Chapter 3, in part, has been submitted for publication of the material as it may appear in the Proceedings of the National Academy of Sciences, 2014, Gneezy, Uri; Imas, Alex O., National Academy of Sciences, 2014. The dissertation author was the co-primary investigator and author of this paper.

Chapter 4, in part, has been submitted for publication of the material as it may appear in Management Science, 2014, Gneezy, Uri; Imas, Alex O.; Madarász, Kristóf, INFORMS, 2014. The dissertation author was the co-primary investigator and author of this paper.

VITA

2007	B. A. in Economics *with Honors*, Northwestern University
2010	M. A. in Economics, University of California, San Diego
2014	Ph. D. in Economics, University of California, San Diego

ABSTRACT OF THE DISSERTATION

Emotional Dynamics in Social Behavior and Choice Under Uncertainty

by

Alex Oleg Imas

Doctor of Philosophy in Economics

University of California, San Diego, 2014

Professor Uri Gneezy, Chair

My dissertation examines how emotions induced by prior outcomes and choices affect preferences, and the ways in which individuals anticipate these behavioral affects and adjust their choices ex-ante.

The first chapter explores a contradiction in the literature documenting how prior losses affect risk attitudes: some studies find that people become more risk seeking after a loss, whereas others find that they become more risk averse. I show that these seemingly inconsistent findings can be explained by individuals' differential responses to realized versus paper losses. Following a realized loss, individuals are sensitized to further losses and avoid risk; if the loss has not been realized – a paper loss – individuals take on greater risk.

The second chapter studies the effectiveness of prosocial incentives, where ef-

fort is tied directly to charitable contributions. I show that people work harder for charity than for themselves, but only when incentive stakes are low. When stakes are raised, prosocial incentives are no longer superior, suggesting that such incentive schemes should only be used when funds are limited. Additionally, individuals correctly anticipate these effects, choosing to work for charity at low incentives and for themselves at high incentives. These results are consistent with a model of warm-glow giving.

The third chapter, co-authored with Uri Gneezy, examines whether individuals use anger strategically in interactions. We first show that in some environments anger helps performance, while in others anger hurts. We then show that individuals anticipate these effects and strategically use the option to anger their opponents in competitions. This finding suggests people understand the effects of emotions on behavior and exploit.

The final chapter, co-authored with Uri Gneezy and Kristóf Madarász, presents theory and experiments where people's prosocial attitudes fluctuate over time following norm violations. We show that people who made an immoral choice were more likely to donate to charity, and that immoral behavior increases when people are informed of a future donation opportunity. We propose a model where guilt induced by past immoral actions increases prosocial behavior, and discuss its importance in charitable giving and in the identification of norms in choice behavior through time-inconsistency.

Chapter 1

The Realization Effect: Risk-Taking After Realized Versus Paper Losses

> *"A person who has not made peace with his losses is likely to accept gambles that would be unacceptable to him otherwise."*
>
> – Kahneman and Tversky (1979)

> *"Losses that come on the heels of prior losses may be more painful than average. Conversely, after a prior loss, the person becomes more loss averse."*
>
> – Barberis, Huang and Santos (2001)

1.1 Introduction

Understanding how prior outcomes affect risk attitudes is critical for the study of choice under uncertainty. Standard expected utility theory assumes prior outcomes can only influence risk-taking if there is a substantial change in wealth (Savage, 1954). In turn, an individual's risk preferences should be stable with respect to small and moderate losses (Rabin, 2000).

However, empirical evidence suggests that risk attitudes are not always independent of the history. Prior losses significantly affect subsequent risk-taking in a variety of settings, from choices over lotteries in laboratory experiments (Thaler and Johnson, 1990) to trading decisions of experienced market-makers (Coval and Shumway, 2005; Liu et al., 2010) and investors (Kaustia and Knupfer, 2008). Research in this domain has produced contradictory empirical results: following a loss, individuals have been shown to become either more risk seeking (Andrade and Iyer, 2009; Weber and Zuchel, 2005; Langer and Weber, 2008), or more risk averse (Shiv et al., 2005; Liu et al., 2010). The empirical contradiction has produced a similar contradiction in theoretical work on dynamic choice under uncertainty. For example, in the models of Barberis et al. (2001) and Dillenberger and Rozen (2014) individuals respond to losses by taking on less risk, while in Shefrin and Statman (1985) and Weber and Camerer (1998) losses lead to more risk-taking. The extent of this inconsistency is incapsulated by the two quotations above, both from papers exploring how prior losses affect risk attitudes. The first statement suggests that individuals take on *more* risk after a loss, whereas the second posits the opposite, that a loss leads to *less* risk-taking.

In this paper, we attempt to reconcile these seemingly inconsistent findings within a unified framework. In doing so, we draw a distinction between a loss that is realized (e.g. selling a losing stock or cashing out after a loss) and one that is not realized – a paper loss (e.g. holding a losing stock or not cashing out after a loss). Using existing data and new experimental evidence, we demonstrate that individuals take on *more* risk after a paper loss and *less* risk after a realized loss. Our first experiment provides support for the main predictions: *ceteris paribus*, the same loss is followed by less risk-taking if it is realized and by more risk-taking if it is not. Our second experiment replicates these results in an environment akin to a casino – where risk-taking is detrimental to wealth – and shows that giving individuals flexibility in realizing their asset positions leads to greater loss chasing and lower earnings than when realization is imposed exogenously.

Our proposed distinction in how individuals respond to realized versus paper losses has important implications for commitment and monitoring in contexts where loss chasing bears negative consequences for wealth (e.g. casino gambling). Particularly, an individual with flexibility in realizing his wealth position will want to avoid experiencing negative realization when he is in the red – a phenomenon known as the *disposition effect* (Shefrin and Statman, 1985; Odean, 1998). Unrealized prior losses lead to even greater risk-taking as the individual attempts to climb out of the "hole." Given such loss chasing, flexibility in realizing one's position may lead to greater losses than if realization was imposed exogenously, particularly in contexts where continued risk-taking may be detrimental to wealth.[1] Section 3 provides direct evidence for this behavioral pattern.

The results of this paper offer a unifying principle – realization – that reconciles two rich strands of literature on the dynamics of risk attitudes (Barberis, 2013; Barberis et al., 2001; Benartzi and Thaler, 1995; Gneezy and Potters, 1997). Particularly, the differential effect of realized versus paper losses on risk-taking contributes to the emerging literature on how non-standard factors such as emotions (Caplin and Leahy, 2001, 2004; Caplin, 2003; Koszegi, 2006) and other psychological factors (Loewenstein, 1996; Loewenstein et al., 2001) affect risk attitudes.

In Section 4, we outline a potential theoretical framework that produces the observed behavioral pattern. The main result follows from one simple assumption: paper losses are integrated and evaluated jointly with prospects in the same choice bracket while realized losses are not. In this framework, individuals engage in *narrow framing* – they evaluate choices and gambles within a given mental account or choice bracket separately from other sources of wealth (Thaler, 1985; Rabin and Weizsäcker, 2009). Choices within a bracket affect each other and are evaluated jointly (Read et al., 1999), while choices in different brackets are evaluated independently.[2] A paper loss is integrated into the decision problem, and prospects are evaluated jointly with respect to this loss. In turn, a loss-averse individual becomes *more* willing to take a gamble if it offers a possibility of erasing the previous negative outcome. Realization stops the integration process and updates the reference point: the individual closes the choice bracket of prior outcomes, internalizes the loss and evaluates subsequent prospects relative to a new choice bracket. In turn, this framework predicts that

[1] Rick and Loewenstein (2008) discuss contexts in which individuals may engage in suboptimal decision-making to recoup losses when they perceive themselves as "in the hole."

[2] For example, betters on the racetrack are far more likely to favor the longshot at the end of the day if they had suffered earlier losses (Ali, 1977), presumably because the betting day serves as a distinct mental account or 'choice bracket': if betting choices were integrated with the rest of the individual's wealth, then prior losses would be too trivial to affect subsequent behavior. For further evidence of choice bracketing, see Barberis et al. (2006); Kahneman and Lovallo (1993).

individuals should take on more risk after a paper loss than a realized one.

The paper proceeds as follows. Section 2 reviews prior work and provides empirical support for the differential effect of realized versus paper losses on risk-taking. Section 3 presents two investment experiments which provide direct support for our predictions. Section 4 outlines a potential framework consistent with the observed results. Section 5 concludes with a discussion of implications for optimal monitoring and contracts.

1.2 An Empirical Contradiction

In this section we analyze existing empirical evidence and demonstrate that distinguishing between realized and paper losses reconciles the contradictory results. We begin with studies conducted in the lab, and then discuss non-experimental data.

Langer and Weber (2008) adapted the investment game of Gneezy and Potters (1997) to study risk-taking in a dynamic context. In the game, individuals were asked to make a series of investment decisions from an initial endowment. Over the course of 30 rounds, participants could either invest part of their endowment in a risky, positive expected-value asset, or to keep it. After each round, a randomization device determined whether the investment would be lost or multiplied. Importantly, if the investment was lost, the participant learned this information but did not part with the loss. All earnings were realized at the end of the experiment. If the participant ended the study with less money than the endowment, he would pay the difference to the experimenter; if he ended the study with more money than the endowment, the difference would be paid to him. In turn, participants in this study experienced *paper* gains and losses after each round prior to the end of the experiment.

We obtained and analyzed the individual-level data from the authors to examine risk-taking following a gain or a loss. As can be seen in Figure 1.1a, overall investment in risk increased as the rounds progressed. We separated the data by investment after a loss and investment after a gain, and calculated the change in investment after each outcome. Running a non-parametric Mann-Whitney test on the results reveals that the increase in risk-taking was driven by participants' responses to losses: individuals took on significantly *more* risk after a loss relative to both a prior gain ($p < .001$) and the benchmark of no investment change ($p < .001$).

Shiv et al. (2005) used a similar investment game to explore the effects of prior gains and losses on risk attitudes. As in Langer and Weber (2008), individuals started the experiment with an endowment and made a series of investment decisions over the course of 20 rounds. In each round, participants made the choice of either investing part of the endowment in a risky, positive expected-value asset, or keeping it. After each round, a randomization device determined whether the investment would be lost or multiplied. Unlike in Langer and Weber (2008), however, earnings were realized after *every single round*. After each round, if the participant learned the investment was lost, he took that part of his endowment and gave it to the experimenter; if the investment was multiplied, the participant received the positive difference. As such, participants in Shiv et al. (2005) experienced *realized* gains and losses after each round.

Figure 1.1b shows investment in risk over time from Shiv et al. (2005). Unlike in Langer and Weber (2008), overall risk-taking decreased as the rounds progress. A similar analysis to the one above reveals that changes in investment are again driven by a differential response to losses, but in the opposite direction. Relative to investment after a prior gain, participants responded to a loss by taking on *less* risk ($p < .01$), and did not change their investment behavior in response to gains.

A review of the literature reveals that the distinction between realized and paper losses reconciles the contradictory findings. Andrade and Iyer (2009) gave participants an endowment of \$10 and asked them to make two rounds of betting decisions on a gamble akin to a roulette wheel. Participants could bet up to \$5 in each around. As in Langer and Weber (2008), earnings were realized at the end of the study, such that first round outcomes were paper gains or losses. The authors found that participants took on significantly *more* risk after a loss (Experiment 2). Using a similar design of sequential investment decisions with paper outcomes, Barkan and Busemeyer (1999) and Weber and Zuchel (2005) found similar results: individuals took on more risk after a loss. Alternatively, Shiv et al. (2006) found that individuals took on less risk after a loss; in their experiment, outcomes were realized after every round.[3]

[3]It is important to note that in *all* studies we examine, taking on risk can erase a prior loss if

Non-experimental studies examining the dynamics of risk attitudes present a similar contradiction. Coval and Shumway (2005) and Liu et al. (2010) studied the risk-taking behavior of professional traders. Both papers found that morning losses significantly affected risk-taking in the afternoon. Such behavior is not consistent with expected utility theory since, given the horizon of one day, wealth effects from prior losses were negligible and agency concerns were neutral. However, Coval and Shumway (2005) found that morning losses led traders to take on more risk in the afternoon, while Liu et al. (2010) found the opposite – morning losses led to less risk-taking in the afternoon. When discussing discrepancies between the two papers, Liu et al. (2010) note a difference in the composition of paper and realized losses. A larger portion of morning losses were realized in Liu et al. (2010) than in Coval and Shumway (2005). In light of our experimental findings, the difference in the ratio of realized to paper losses may explain the contrast in the two sets of results.

Although these studies provide suggestive evidence for the differential effect of realized versus paper losses on risk-taking, it is not causal. Besides differences in the nature of losses, there were several other discrepancies which may have contributed to the contrasting results: participants were drawn from different subject pools or populations, the risky assets had different distributions, etc. We now proceed to demonstrate the proposed behavioral pattern within the same experimental paradigm.

1.3 Two Investment Experiments

To test whether paper versus realized losses have differential effects on subsequent risk attitudes, we adopted the investment game of Gneezy and Potters (1997) to examine how the nature of a prior loss affected individuals' risk-taking in a sequence of investment decisions. The decision maker receives an endowment, $\$E$, and makes investment decisions over a series of rounds. In each round, he can choose how much of an amount, X, he would like to invest in a risky option and how much to keep. The amount invested in the risky option, x, yields a dividend of kx $(k > 1)$

the gamble or investment is successful. As discussed in Section 4, we argue that this is a critical motivation for increased risk-taking after a paper loss. For example, Thaler and Johnson (1990) and Heath (1995) do not find support for loss chasing if the gamble does not allow the individual to offset the prior loss.

with probability p and is lost with probability $1-p$. The money not invested $(X-x)$ is kept by the decision maker. The payoff in each round is:

$$p \cdot (X - x + kx) + (1 - p) \cdot (X - x)$$

After the choice of x is made, the outcome of the risky option is determined and revealed to the decision maker. The decision maker then moves on to the next round where he is presented with the same choice.

The amount invested x provides a robust metric for capturing treatment effects and differences in attitudes toward risk. Similar paradigms have been used to test for myopic loss aversion in students (Gneezy and Potters, 1997) and professional traders (Haigh and List, 2005), to demonstrate decreased (and increased) risk-taking following a prior loss (Shiv et al., 2005; Langer and Weber, 2008), to examine gender differences in risk attitudes (Charness and Gneezy, 2012) and to show the effect ambiguity aversion and illusion of control on portfolio choice (Charness and Gneezy, 2010).

1.3.1 Realized and Paper Losses

Undergraduates (N=129) from a university-wide subject pool were recruited to participate in an experiment on decision-making. Participants were randomly assigned to individual computer stations and given a set of instructions that were read aloud. Each was endowed with an envelope of \$8 in cash in the beginning of the study and asked to count it: the envelope contained 7 one-dollar bills and 4 quarters.

Participants were told that they would make 4 rounds of investment decisions. In each round the participant would decide how much of \$2 to invest in a lottery (in increments of quarters). With a 1/6 chance the lottery would succeed and pay dividends $k = 7$ times the amount invested x; with a chance of 5/6 the lottery would fail and the money invested would be lost. In each round, participants were randomly assigned one "success number" between 1 and 6. This number was displayed on their computer screen in the beginning of each round. Participants would then enter the amount they would like to invest x. Note that in this case, p (1/6) and k (7) were chosen such that that $p \cdot k > 1$, making the expected value of investing higher than the expected value of not investing.

Once this was done, the experimenter rolled a six-sided die in the front of the room. Participants were welcomed to examine the die to make sure it was fair. If the outcome of the die roll matched a participants success number, the lottery would succeed and they would earn $7x$ plus the amount they did not invest $(2 - x)$. If the outcome was any other number, the lottery would fail and participants were left with the amount they did not invest. After learning the outcome of the die roll, participants would move on to the next round, be assigned a new "success number" and make the same decision again. All outcomes were written on a board in front of the room to keep information constant between treatments.

To test for the differential effect of realized versus paper losses, participants were randomized into either the Realized or Paper treatment. In the Realized treatment, at the end of the 3^{rd} round participants had their wealth positions realized: if they had lost money by the end of the 3^{rd} round, they took this amount out of their envelope and handed it to the experimenter. If they had won, this amount was given to them. After realizing their earnings, participants made one last investment decision in the 4^{th} round and were paid according to the outcome.

In the Paper treatment, participants did not realize their earnings at the end of the 3^{rd} round. They continued on to the 4^{th} round and were paid at the end of the experiment. Time between rounds was normalized across treatments such that those in the Realized treatment did not have a longer break between the 3^{rd} and 4^{th} rounds than those in the Paper treatment. As a robustness check to ensure that the intervention of the experimenter in the Realized treatment did not drive the results, we also ran a second Paper Social (Paper S) treatment. The procedure in Paper S was the same as in the Paper treatment, except that at the end of the 3^{rd} round the experimenter came up to each participant and verbally informed them how much money they had won or lost relative to the original endowment.[4]

Note that given a sequence of decisions and outcomes, the wealth positions and information were the same for participants in all three treatments. However, those who had lost money from their \$8 endowment by the end of the 3^{rd} round in the

[4] We ran the Paper S treatment as a robustness check after running the Paper and Realized treatments, using the same lab and subject pool.

Realized treatment had to physically part with it. This served as the manipulation of exogenously inducing realization. In contrast, those who had a similar loss by the end of the 3^{rd} round in the Paper and Paper S treatment could still potentially avoid parting with their endowment and experiencing a negative realization by taking on more risk in the 4^{th} round.

In this setup, we predict that those who are losing at the end of the 3^{rd} round in the Paper and Paper S treatments should *increase* their 4^{th} round investment. In contrast, those who are losing at the end of the 3^{rd} round in the Realized treatment should *decrease* their investment in the 4^{th} round. In the analysis below, we use the change in investment between rounds as our main dependent variable.

Results

Investments did not significantly differ by treatment in rounds 1 through 3 (see Table 1.1). However, participants in the Paper and Paper S treatment took on significantly more risk in the 4^{th} round (average investment $\bar{x}$ = \$1.01 and \$.96, respectively) than those in the Realized treatment ($\bar{x}$ = \$0.63; $t(79) = 2.71, p < .01$ and $t(86) = 2.38, p = .02$).[5] In line with our predictions, the variation between the two treatments was largely driven by differences in how individuals responded to a loss that was realized versus one that was not.

We examine the investment change between rounds 3 and 4 for participants who lost the lottery in rounds 1, 2 and 3 – the lottery failed all three times (see Figure 1.2). Consistent with our predictions, in the Paper and Paper S treatments these individuals increased their investment in the lottery by \$0.23 and \$0.16, respectively, taking on significantly *more* risk than the null hypothesis of zero change in investment ($t(26) = 2.28, p = .03$ and $t(38) = 2.41, p = .02$).[6] However, those who had similarly lost in the Realized treatment decreased their investment by \$0.15, taking on significantly *less* risk than the null ($t(25) = 2.42, p = .02$). Importantly, the change in investment between the 3^{rd} and 4^{th} rounds was significantly different between both the Paper and Realized treatments ($t(51) = 3.19, p < .01$) and the Paper S and Realized treatments ($t(63) = 3.25, p < .01$). To quantify the effect of realizing

[5]Unpaired, two-sample t-test.

[6]One-sample t-test with null hypothesis of zero investment change.

a loss on risk-taking, an OLS regression of Investment Change on a treatment dummy (Realized = 1; Paper = 0) for those who lost by the end of the 3^{rd} round revealed that realizing one's loss leads to a significant decrease in risk-taking relative to not realizing a loss of the same size ($\beta = -.38, p < .01$). Changes in investment did not differ between treatments for any other round (see Appendix Table 3).[7]

Next, we examine the relationship between the amount lost and investment behavior. Looking at each treatment separately, we regress the amount invested in the 4^{th} round on earnings at the end of the 3^{rd} round relative to earnings if the individual had not chosen to gamble ($6), e.g. someone who had lost $3 by the 4^{th} round would have relative earnings of −$3. Relative earnings had a significant effect on 4^{th} round investment if the outcome was not realized ($\beta = -.09, p < .01$). Note that the negative coefficient implies that the larger the losses, the more the individual invested in the 4^{th} round. In contrast, relative earnings had no effect on investment if outcomes were realized ($\beta = -.002, p = .92$). Regressing the amount invested on relative earnings, a dummy for realization and their interaction reveals a significant interaction effect ($\beta = -.09, p = .015$). This implies that relative earnings had a negative effect on the amount invested, but only if the outcome was not realized.

These results are consistent with the framework outlined in Section 4, where paper prior outcomes are integrated within the same choice bracket as the prospect being evaluated, while realized outcomes are not. A loss averse individual with prospect theory preferences would invest more the greater his losses relative to the reference point – the more he is in the hole. However, this only occurs if the prior losses are integrated with the prospect; if the losses are not integrated, i.e. realized outcomes, the extent of the loss should not affect investment behavior. This pattern is consistent with our results.

1.3.2 Realization and Flexibility

The second experiment aimed to explore whether giving individuals flexibility in when to realize their positions could lead to lower earnings overall relative to

[7]Since the experiment was designed to test the effect of prior losses, probabilities were chosen such that most participants had lost by the end of the 3^{rd} round. There were too few observations to conduct any meaningful analyses on prior gains.

those whose positions were exogenously realized. Particularly, the experiment was designed to mimic environments where taking on risk and chasing losses leads to lower expected returns than keeping one's money (e.g. casinos, race-tracks), and the choice to realize one's position is endogenous. The disposition effect predicts that most would choose to not realize their losses, and in light of our findings, take on greater risk. On the other hand, imposing realization exogenously should mitigate this behavior. Hence, flexibility in realization is predicted to lead to lower expected wealth due to a combination of loss chasing and the disposition effect.

Undergraduates (N=150) from a university-wide subject pool were recruited to participate in an experiment on decision-making. The lottery was set to yield 2.5 times the amount invested $x with a 1/3 probability, and to lose with probability 2/3. Since $p \cdot k < 1$, the expected value of investing in the lottery is slightly lower than the expected value of not investing, similar to gambling on a roulette wheel. Procedures were largely the same as those in the first experiment, except that now participants were randomly given two different "success numbers" from 1 to 6 at the beginning of each round to reflect the higher probability of the lottery succeeding.

In addition to the Paper and Realized loss treatments, a third Flexible treatment was added to test whether flexibility in realization indeed reduced expected earnings. In the Flexible treatment, individuals were asked at the end of the 3^{rd} round whether they would like to realize their earnings similar to those in the Realized treatment, or to move on to the 4^{th} round. If they chose to realize their positions, the procedure was identical to the Realized treatment; if they chose to move on, the procedure was identical to the Paper treatment.

Similar to the first experiment, we predict that those who had a paper loss by the end of the 3^{rd} round would increase their position in the lottery, taking on more risk, while those who had a realized loss would decrease their position and take on less risk. Moreover, we predict a disposition effect in the Flexible treatment – those who had won by the end of the 3^{rd} round should want to realize their positions to a greater extent than those who had lost, with the latter group preferring to move on to the 4^{th} round without realization. In turn, participants who had lost by the end of the 3^{rd} round in the Flexible treatment should increase their position in the risky lottery in the 4^{th} round since most had not realized their loss.

Results

Investments did not significantly differ by treatment in rounds 1 through 3 (see Table 1.2). However, participants in both the Paper treatment ($\bar{x} = \$1.16$) and Flexible treatment ($\bar{x} = \$1.24$) invested significantly more in the 4^{th} round than those in the Realized treatment ($\bar{x} = \$0.74$; $t(78) = 2.68, p < .01$ and $t(79) = 3.49, p < .001$, respectively).[8] Again, these differences were largely driven by the differential response to prior losses that were realized versus those that were not.

In line with our predictions (see Figure 1.3), those who had lost by the end of the 3^{rd} round in the Paper treatment on average increased their investment in the lottery by \$0.29, taking on significantly *more* risk than the null hypothesis of zero investment change ($t(25) = 2.75, p = .01$).[9] Similarly, those in the Flexible treatment who lost by the end of the 3^{rd} round *also* took on more risk, increasing their investment in the lottery by \$0.33 ($t(23) = 3.14, p < .01$). In contrast, those who had lost by the end of the 3^{rd} round in the Realized treatment took on *less* risk and decreased their subsequent investment by \$0.27 ($t(25) = 2.19, p = .038$). Investment changes after a loss between the 3^{rd} and 4^{th} rounds were significantly different between the Realized and Flexible treatments ($t(48) = 3.68, p < .001$) and the Realized and Paper treatments ($t(50) = 3.44, p = .001$), but not between the Paper and Flexible treatments ($t(48) = .37, p = .71$). Changes in investment did not differ between treatments for any other round (see Appendix Table 4).

Importantly, participants in the Flexible treatment displayed a significant disposition effect. Participants who had lost by the end of the 3^{rd} round realized their positions 13% of the time while those who had won realized their positions 44% of the time ($t(38) = 2.33, p = .025$). Since most who lost by the end of 3^{rd} round chose not to realize those losses, our predictions state that participants in the Flexible treatment should take on more risk in the 4^{th} round, which was observed in the results.

Greater loss chasing had a significant negative impact on earnings. Given that investing in the lottery had a lower expected return than not investing, participants in the Paper and Flexible treatments stood to earn less in the 4^{th} round than those for whom realization was imposed exogenously. Indeed, expected 4^{th} round earnings in

[8] Unpaired, two-sample t-test.

[9] One-sample t-test with null hypothesis of zero investment change.

the Realized treatment were significantly higher than in both the Paper and Flexible treatments ($t(78) = 2.68, p < .01$ and $t(79) = 3.49, p < .001$, respectively). Hence, giving individuals flexibility in realizing their positions in a context where risk-taking is detrimental to wealth led to greater losses than when realization was imposed exogenously.

1.4 Discussion

In one of the first papers to make a distinction between realized and paper losses, Shefrin and Statman (1985) posit that when an individual buys a stock or makes an investment, a mental account is opened. If the asset is sold for lower (higher) than the initial purchase price, the individual codes this as a realized loss (gain). Realizing a gain or loss closes the associated mental account. On the other hand, if the asset goes down (up) before it is sold, the individual codes this as a paper loss (gain). Individuals experience a pronounced drop in utility when a mental account is closed at a loss; the authors argue that a realized loss is more painful than a paper one. Barberis and Xiong (2012) formalize the distinction between paper and realized outcomes in a model of *realization utility*, which corresponds to an individual's feelings of regret (elation) upon closing his mental account with a wealth position lower (higher) than when he started (see Ingersoll and Jin (2013) for an extension of these results).[10]

A natural question is whether agents' risk-taking differs after a loss prior to realization (a paper loss) versus after a loss that is realized.. Prospect theory (Kahneman and Tversky, 1979; Wakker and Tversky, 1993) has been used to explain many documented instances of increased risk-taking following a prior loss (Weber and Camerer, 1998; Coval and Shumway, 2005; Smith et al., 2009; Weber and Zuchel, 2005). When discussing dynamic effects of prior outcomes, Kahneman and Tversky (1979) observe that a "person who has not made peace with his losses is likely to accept gambles that would be unacceptable to him otherwise." Particularly, if taking on risk allows an individual to offset a prior loss and avoid a negative realization,

[10] A recent study by Frydman et al. (2012) provides neural support for realization utility: activation in brain regions associated with expected utility was more significant after a realized outcome.

then he will take on more risk than if no prior loss had occurred.

Critical to this prediction is the assumption that people *integrate* the outcomes of successive gambles when evaluating prospects. As noted in Thaler and Johnson (1990), if a prior loss is not integrated with the prospect, subsequent risk-taking may lead individuals to take on less risk. Setting conditions for when a decision maker integrates prior outcomes and when he does not – evaluating prospects with a "clean slate" – is key for determining when a prior loss leads to more risk-taking and when it may lead to less. Note that not integrating a prior outcome is analogous to incorporating it into total wealth and updating (resetting) the reference point to the new wealth level.

This paper posits that integration depends on whether the prior outcome was realized or not: individuals integrate prior paper losses and update their reference point after realization. After a paper loss, the individual may feel that there is still hope he can still recoup his loss and avoid the sure negative realization – he has not "made peace" with his losses. When evaluating prospects, the losses are integrated with the potential payoffs, and as such, gambles that allow the individual to erase the prior outcomes become more attractive. On the other hand, the realization of a loss serves as a natural point for an individual to internalize the negative outcome and close the associated choice bracket (Shefrin and Statman, 1985). Particularly, if a loss is realized – the individual parts with the money – he does not integrate the prior outcome when evaluating a prospect and updates his reference point. Conditional on the gamble allowing the individual to recover from the prior loss, he should be less likely to take on risk after a realized than a paper loss. Consistent with the empirical results, loss chasing should only be observed after a paper loss.

As a simple illustration, take decision maker with with prospect theory preferences and a piecewise linear value function.[11] Suppose he evaluates gains and losses relative to a reference point of zero and has a loss aversion parameter $\lambda = \frac{5}{4}$.[12] Such a decision maker will be indifferent between taking or leaving a gamble with a one

[11]See the Appendix for a more general illustration of the framework.

[12]Kahneman and Tversky (1979) and Thaler and Johnson (1990) take the reference point r to be the status quo. Koszegi and Rabin (2006) formulate the value function as $v(m(x) - m(r))$, where the reference point r is characterized by the full distribution of expected outcomes and $m(\cdot)$ is the "consumption utility" typically studied in economics. Here, we take r to be the status quo for simplicity; all results hold if r is defined in terms of expectations.

third chance of winning \$250 and a two thirds chance of losing \$100.

Suppose the decision maker takes the gamble and suffers a \$100 paper loss. If offered the same gamble again, he now compares the sure \$100 loss if he turns it down to the prospect of either winning \$150 or losing \$200 if he accepts the second gamble. Since the gamble allows the decision maker to avoid realizing the sure loss, it is straightforward to show that he now strictly prefers to take the gamble. On the other hand, if the \$100 loss was realized, the prospective gamble is evaluated in a new choice bracket and the decision maker does not integrate the two when choosing to accept or reject the second gamble. As such, he strictly prefers to take the gamble after a paper loss, but is indifferent after a realized loss.

In order to show that individuals take on less risk after a realized loss not only relative to a paper loss, but relative to a baseline with no prior outcomes, more structure is needed. An account proposed in the literature is that losses which are not integrated with prospects may cause the individual to become more loss averse (Thaler and Johnson, 1990; Barberis et al., 2001). The increased loss aversion may be due to a diminished capacity for dealing with bad "news" about future consumption (Koszegi and Rabin, 2009; Pagel, 2012) or the increased salience of the potential downside of risk (Bordalo et al., 2012). In the framework of the above example, we assume that a realized loss is not integrated with the prospect. If it is further posited that the loss increases subsequent loss aversion, then the decision maker strictly prefers not to gamble after a realized loss – he is less willing to take on the second gamble than the first.

It should be stressed that the preceding arguments are meant as a preliminary framework to rationalize the observed differential effect of realized versus paper losses on risk-taking. Future work is needed to both identify the mechanism behind the phenomenon and develop a general theoretical structure.

1.5 Conclusion

In this paper we present the results of two experiments which demonstrate the differential effect of paper versus realized losses on risk attitudes. Individuals take on *more* risk after a paper loss and *less* if the loss is realized. Using a framework of prospect theory and choice bracketing, we argue that the realization of a loss acts as a closing of the choice bracket constituting the individual's prior investments or gambles. If the loss is not yet realized, however, people attempt to cover it by taking on even more risk – chasing their loses – in order to avoid a future negative realization. Importantly, we also demonstrate that flexibility in realizing investment decisions may be detrimental to wealth (in certain contexts) due to escalations in risk-taking.

The interplay between loss chasing and the disposition effect has significant implications for the role of choice and monitoring of investor behavior. The results of our second experiment demonstrated that individuals whose investments were unsuccessful were reluctant to realize their losses, preferring to instead take on more risk before their positions were finally realized. These effects can be particularly detrimental in contexts where taking on gambles or investing in particular assets leads to lower expected returns than available alternatives.

Given the behavioral patterns described above, an individual owning a losing stock may be reluctant to sell, choosing to instead keep the stock or even "doubling down" and purchasing more shares. Such trading behavior can spiral out of control and lead to significant losses, which is consistent with the literature documenting overly aggressive trading (Barber et al., 2006) and a pronounced disposition effect displayed by individual investors (Odean, 1998). The results presented here suggest that giving traders flexibility in realizing their asset positions may lead to significantly lower wealth. In turn, individual investors should adopt policies which facilitate regular realization to prevent detrimental loss chasing. For example, an individual can automatically set his asset positions to be reported to a third party who can exogenously influence the realization of his positions. Our findings are also related to the prescriptions proposed by Weber and Zuchel (2005) regarding the benefits of binding precommitments in investment strategies.

For institutional traders, given the relationship between compensation and trading performance, the reporting of asset positions to the overseeing risk manager can be taken as a natural point of realization – analogous to the closing of the respective choice bracket. Anecdotal evidence suggests that some of the largest losses suffered by financial institutions occurred as a result of traders hiding prior losses while taking on excessive risk in an attempt to cover them.[13] In light of our results, a firm's monitoring strategy should utilize realization of traders' asset positions while lowering the incentives for them to hide loses. For example, Camerer and Loewenstein (2004) describe an investment banker whose firm forced traders to periodically switch positions (the portfolio of assets that the trader bought and is blamed or credited for) with the position of another trader. This was done to ensure that traders do not make bad trades because of emotional attachment to their previous actions, while keeping the firm's net position unchanged. In the context of our findings, such a policy would be an effective tool to curb loss chasing, particularly if position switches were performed at times not announced to the traders ex-ante.

Future research should extend our findings to field settings and explore the role of realization in mitigating commonly reported pitfalls in investment and gambling behavior. Particularly, it is important to understand to what extent individuals are aware of the differential effects of realized versus paper losses on their future risk-taking behavior, and whether they would value opportunities and commitment devices that would allow them to curb detrimental loss chasing.

1.6 Appendix

1.6.1 Choice Bracketing and Realization

Consider a dynamic choice problem where at each time $t \in \mathscr{T} = \{1, ..., T\}$, a decision maker faces a choice set $\{L_t, r_t\}$ which represents a decision between taking a non-degenerate, positive expected-value gamble L_t or turning it down and retaining r_t, where r_t corresponds to the reference point at the start of period t. L_t has two possible outcomes, $x^g > 0 > x^l$, where x^s is the payout in state $s \in \{g, l\}$. Denote the objective probability that state g occurs by π, with corresponding probability $1 - \pi$ for state l.

A loss-averse decision maker evaluates outcomes using the prospect theory value function $v(x)$. In turn, the decision maker evaluates the gamble L_t as $V(L_t) =$

[13] See Jerome Kerviel's 4.9 billion Euro loss for Societe Generale, Kweku Adoboli's 2.3 billion dollar loss for UBS and Nick Leeson's 1.3 billion dollar loss for Barrings, which wiped out the firm.

$\pi^g v(x^g) + \pi^l v(x^l)$. Lottery outcomes are evaluated relative to a reference point of zero, $r_t = 0$.

To examine the differential effects of realized versus paper outcomes on risk-taking, we first define the role of realization in determining whether outcomes are evaluated jointly within the same choice bracket– integration – or separately in different choice brackets.

Let $\{R_t\}_{t=2}^{T}$ be the stochastic process indicating whether a realization of an outcome – selling of a stock or parting with money lost in a gamble – occurs at time t, $R_t : s \times \mathscr{T} \rightarrow \{0, 1\}$. Let $\{\tau_n\}$ be a sequence of hitting times ($R_{\tau_n} = 1$) such that $\tau_1 = inf\{t \in \mathscr{T} \ s.t. \ R_t = 1\}$ and $\tau_n = inf\{t \in \{\tau_{n-1} + 1, ..., T\} \ s.t. \ R_t = 1\}$ for $n \geq 2$.[14] Then the decision maker's value function can be defined as:

$$v(x_t, \lambda_t) = \begin{cases} (x_t + \sum_{s}^{t-1} L_s)^\alpha & if \ v(\cdot) \geq 0 \\ -\lambda(z_t) \cdot [-(x_t + \sum_{s}^{t-1} L_s)]^\alpha & if \ v(\cdot) < 0 \end{cases} \tag{1.1}$$

where $s = max\{\tau_n \ s.t. \ \tau_n \leq t\}$. A note on the timing notation: the realization of a prospect evaluated at time t occurs at the start of the next period $t+1$: $R_{t+1} = 1$ if the outcome is realized or $R_{t+1} = 0$ if it is not.

The expression in (1) formally denotes the distinction between paper and realized outcomes. Particularly, the decision maker evaluates a prospect L_t jointly with the preceding sequence of outcomes $\{x_\tau, ..., x_{t-1}\}$ within the same choice bracket if no prior outcome in that sequence had been realized ($\forall \ n \in \{\tau+1, ..., t\}$: $R_n = 0$). The realization of a prior outcome (e.g. $R_t = 1$) closes the choice bracket containing the preceding sequence, and the prospect L_t is evaluated separately in a new choice bracket.

In the last period, the outcome from the final choice is always realized, $R_T = 1$. This feature is analogous to the end of a trading day, where the choice bracket is closed exogenously. In addition, we follow Barberis and Xiong (2012) in assuming that the decision maker derives utility from an outcome according to expression (1) only if the outcome was realized. The strong assumption that individuals only derive *realization*

[14] For example, if $t = 5$ and outcomes had been realized in $t = 3, 5$, then $R_t = \{R_1 = 0, R_2 = 0, R_3 = 1, R_4 = 0, R_5 = 0\}$ and $\tau_n = \{\tau_1 = 3, \tau_2 = 5\}$.

utility is made for simplicity; all results hold if we assumed that the utility derived from a realized outcome was greater than from a paper one.[15]

1.6.2 Dynamics

We follow Barberis et al. (2001) to allow loss aversion λ_t to depend on prior outcomes represented by z_t and assume this dependence takes a linear form $\lambda(z_t) = \lambda \cdot z_t$, with $\lambda \geq 1$.[16] The authors posit that a prior loss makes the decision maker more sensitive to subsequent losses.[17] We take the baseline $z_t = 1$ if the prior outcome x_{t-1} met expectations (equivalent to no prior outcome), and let $z_t > 1$ if the prior outcome was below expectations (a loss) and $z_t \leq 1$ if the outcome was above expectations (a gain). In turn, the decision maker appears more loss averse after a prior loss, a process termed *sensitization.*

Take $\alpha = 1$ for simplicity. We examine the interplay between sensitization and realization in a choice problem with three periods, $t = 1, 2, 3$. Suppose a decision maker is indifferent between gambling or not in period $t = 1$, $\lambda(z_1) = \frac{\pi x^g}{x^l(\pi-1)}$, and breaks the indifference by choosing to gamble. The gamble results in a loss $x_1 = x^l$ that is not realized ($R_2 = 0$). Our first proposition follows:

Proposition 1 (***Paper Losses***)**.** *A decision maker who experiences a realized loss,* $(R_t = 0)$*, will be more likely to take on risk than one who had not experienced a prior loss.*

Proof. The decision maker takes the gamble *if* $\lambda(z_2) < \frac{\pi(x^g+x^l)}{x^l(2\pi-1)}$. Since the loss was not realized, $\lambda(z_2) = \lambda(z_1)$, and $\frac{\pi x^g}{x^l(\pi-1)} = \lambda(z_2) < \frac{\pi(x^g+x^l)}{x^l(2\pi-1)}$, the decision maker takes on *more* risk after a paper loss. Q.E.D.

On the other hand, if the loss $x_1 = x^l$ was realized ($R_2 = 1$), then our framework makes a clear prediction:

---

[15] Shefrin and Statman (1985) make the point that realizing losses at the closing of a mental account hurts more than the equivalent paper loss within a mental account. Similarly, Thaler (1999) writes "one clear intuition is that a realized loss is more painful than a paper loss."

[16] Barberis et al. (2001) set $\lambda_t = \lambda + k(z_t - 1)$ such that setting $k = 0$ reduces the model to the standard form.

[17] Similarly, Thaler and Johnson (1990) suggest that "a prior loss might even sensitize people to subsequent losses of a similar magnitude (p. 657)." Participants in their experiments anticipated that a loss would hurt more after a prior loss than if it had occurred by itself, suggesting sensitization.

Proposition 2 (***Realized Losses***)**.** *A decision maker who experiences a realized loss,* $(R_t = 1)$*, will be less likely to take on risk than one who had not experienced a prior loss.*

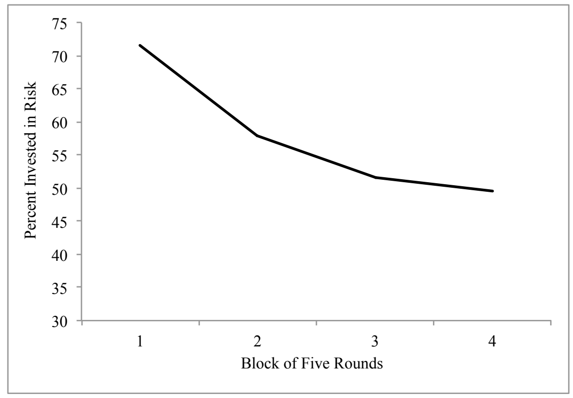
Proof. Realization closes the choice bracket of prior outcomes and the prospect is evaluated independently. Without the motivation of erasing a loss with a possible gain, the decision maker strictly prefers not to take the gamble since $\lambda(z_2) > \lambda(z_1) = \frac{\pi x^g}{x^l(\pi-1)}$. Q.E.D.

Note that in the preceding analysis we have assumed that the decision maker is naive in the sense that he expects to evaluate each prospect with respect to a clean slate. Namely, he does not expect to integrate prior outcomes when evaluating future prospects. In any period t, the decision maker makes choices as if he will use the same value function (1) in $\{t+1, ..., T\}$ as he did in $t = 1$.[18] In our setup, a sophisticated decision maker indifferent between gambling in period $t+1$ will always take the gamble in period t, regardless of the nature of the period t outcome. This prediction is not supported by our experimental results in Section 3.

[18] For evidence on naivitè about future preferences in time discounting, see O'Donoghue and Rabin (1999) and Della Vigna and Malmendier (2006); for naivitè in predicting future tastes, see Loewenstein et al. (2003); for naivitè about future emotions, see Loewenstein (1996).

1.7 Figures

(a) *Langer and Weber* (2008)



(b) *Shiv et al.* (2005)

Figure 1.1: Risk-taking after Realized and Paper Losses

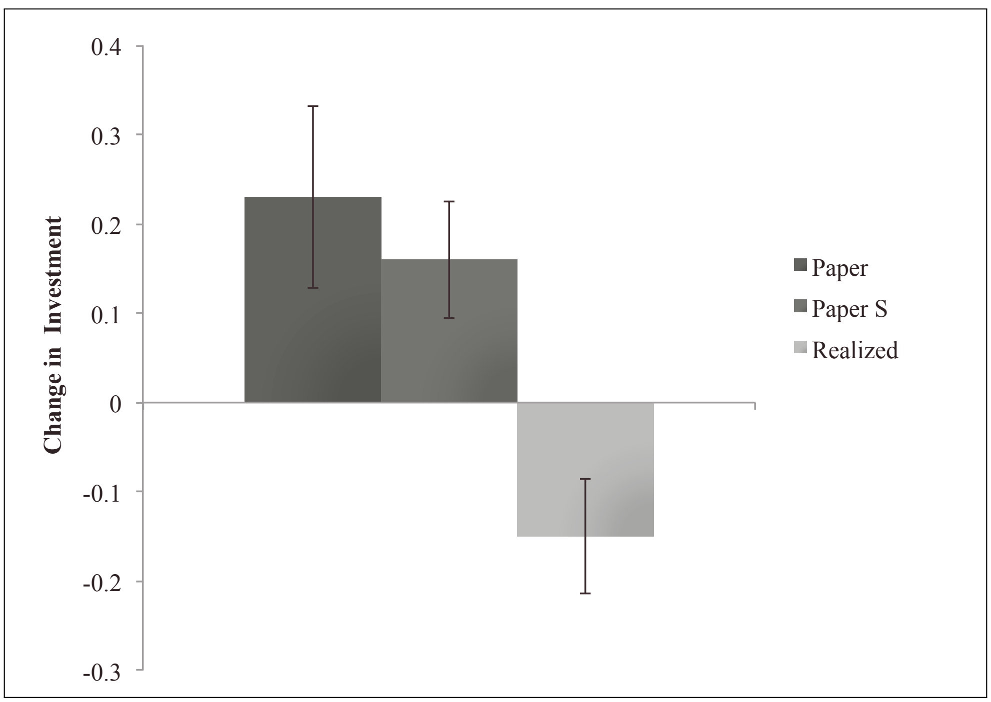


Figure 1.2: Investment Change ($) after a Loss

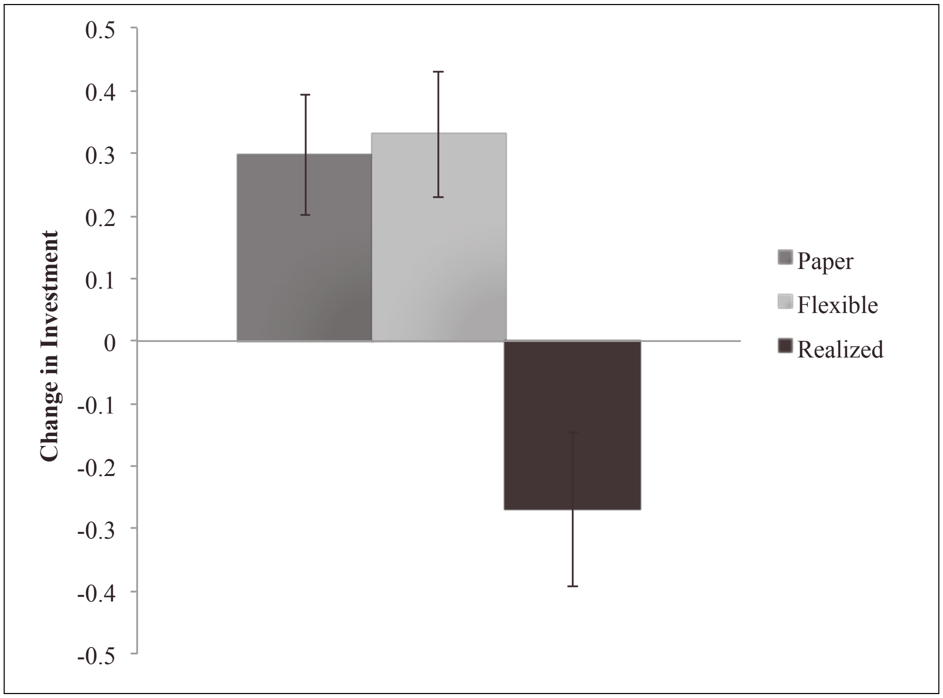


Figure 1.3: Investment Change ($) after a Loss

1.8 Tables

Table 1.1: Total Investment in Risk: Experiment 1

Treatment	Round 1	Round 2	Round 3	Round 4
Realized	$0.82	$0.72	$0.76	$0.63
	(.08)	(.09)	(.09)	(.08)
Paper	$0.85	$0.78	$0.71	$1.01
	(.09)	(.09)	(.10)	(.11)
Paper Social	$0.82	$0.75	$0.76	$0.96
	(.08)	(.08)	(.10)	(.11)

† Standard errors in parentheses.

Table 1.2: Total Investment in Risk: Experiment 2

Treatment	Round 1	Round 2	Round 3	Round 4
Realized	$0.99	$0.96	$0.94	$0.74
	(.09)	(.10)	(.11)	(.10)
Paper	$1.01	$0.94	$0.83	$1.16
	(.09)	(.10)	(.09)	(.11)
Flexible	$1.04	$0.96	$1.06	$1.24
	(.08)	(.10)	(.10)	(.10)

† Standard errors in parentheses.

Table 1.3: Investment Change after Loss at End of Round: Experiment 1

Treatment	Round 1	Round 2	Round 3
Realized	$-0.07	$-0.03	$-0.15
	(.08)	(.08)	(.06)
Paper	$-0.07	$-0.09	$+0.23
	(.06)	(.04)	(.10)
Paper Social	$-0.07	$+0.01	$+0.16
	(.08)	(.07)	(.06)

† Standard errors in parentheses.

Table 1.4: Investment Change after Loss at End of Round: Experiment 2

Treatment	Round 1	Round 2	Round 3
Realized	\$-0.12	\$+0.05	\$-0.27
	(.09)	(.12)	(.12)
Paper	\$-0.09	\$-0.08	\$+0.29
	(.08)	(.11)	(.10)
Flexible	\$-0.10	\$+0.16	\$+0.33
	(.06)	(.12)	(.11)

† Standard errors in parentheses.

Chapter 2

Working for "Warm Glow": On the Benefits and Limits of Prosocial Incentives

2.1 Introduction

Designing incentive schemes to best motivate effort is an important design question for firms, governments and other organizations. Neoclassical economic theory predicts that individuals should excerpt more effort when compensation for performance is paid directly to them rather than to a charity. Money is a good and effort is costly; therefore, indirect compensation through charitable donations or gifts should be a less effective incentive than an equivalent direct monetary payment.

However, evidence from psychology (Deci, 1971) and behavioral economics (Gneezy et al., 2011b) suggests that direct monetary incentives may not always be optimal in motivating performance, especially when the incentive levels are low. Suppose a company wants to motivate its employees to lose weight and can pay at most \$5 per pound lost. Losing a pound is difficult and the extra \$5 in purchasing power may not be motivation enough for individuals to lose the weight.

Alternatively, suppose that the company offered to donate \$5 to a local charity for every pound lost. The pleasure derived from doing good for others every time the treadmill is used or unhealthy food is avoided may be greater than from equivalent direct compensation. In turn, these prosocial incentives could provide sufficient motivation for individuals to take the necessary steps towards weight loss. However, although prosocial incentives may be more effective when the amounts involved are small, would they still work harder for charity when the stakes increase—would people work harder for a $100 or $1000 donation rather than receiving the money directly?

This paper studies whether and when prosocial incentives—where charitable contributions are directly tied to effort levels—lead to better performance and greater effort provision than standard, self-benefiting incentives. We find that individuals indeed provide greater effort under prosocial incentives than under standard, self-benefiting ones when the incentive stakes are low. However, this difference disappears or reverses when the stakes are high. In addition, individuals seem to anticipate these effects and prefer to work for prosocial incentives when the stakes are low and self-benefiting ones when stakes are raised.

In a recent paper, Dunn et al. (2008) demonstrate that individuals report greater happiness when spending money on others than when spending on themselves. Bonuses in the form of charitable contributions were shown to increase employee happiness relative to paying the employee directly (Norton et al., 2012). These findings suggest that, at least over some incentive levels, tying individuals' effort directly to charitable contributions may be a more effective motivator than standard incentives. While numerous lines of research have argued that individuals derive pleasure directly from prosocial behavior (Loewenstein et al., 1989; Charness and Rabin, 2002), the theory of "warm glow" giving offers the most parsimonious framework for testing this intuition (Andreoni, 1988, 1989, 1990). According to the theory, individuals derive private value from the the altruistic *act* apart from the overall outcome provided for others.

In addition, studies have shown that introducing direct monetary incentives may not necessarily increase an individual's motivation to exert effort in a particular task. In some cases, introducing low self-benefiting incentives actually decreases performance relative to providing no incentives at all. Gneezy and Rustichini (2000) showed that, relative to no compensation, students scored worse on an IQ test when they were paid a small amount directly for their performance; Heyman and Ariely (2004) found a similar result in the case of neutral computer task. Drawing on these findings, we posit our first hypothesis (Hypothesis 1): a prosocial incentive scheme may both be preferred and result in better performance than a self-benefiting incentive scheme when the incentive stakes are low.

However, there is reason to suspect that the benefit individuals derive from charitable acts may be to some extent independent of the benefit for the recipients of the acts. In other words, people derive warm glow from giving, valuing the effort exerted for others rather than the benefit others receive from that effort. Andreoni (1993) offered support for this notion through an experimental test of the crowding-out hypothesis, which predicts that government contributions to a privately provided public good (e.g., a charity) should completely crowd out private contributions. The neutrality result presumes that purely altruistic individuals do not place a private value on the act of giving: if the government uses a dollar in taxes for contributions to a charity, then the individual will reduce her voluntary contribution by a dollar while keeping the charity's revenue constant. The theory of warm glow giving makes an alternate prediction—that crowding out will be incomplete since individuals derive a private benefit from the act of personally giving. Andreoni used a modified public goods game to show that indeed crowding out was incomplete: when participants were taxed and taxes were contributed to the public good, they did not reduce voluntary contributions by the size of the tax. Total contributions were significantly larger when participants were taxed than when they were not, suggesting that individuals valued the act of giving apart from the overall impact of contributions.

Hsee and Rottenstreich (2004) provide further evidence of individuals' insensitivity to the benefits of the donation for others, showing that individuals are willing to donate the same amount for the rescue of one animal as for the rescue of four animals. This type of scope insensitivity is also shown by Small et al. (2007), who demonstrate that individuals donate the same amount for one person as for 10 people (also see Linardi and McConnell (2011) for additional evidence of scope insensitivity in the social domain). In light of these studies, our second hypothesis (Hypothesis 2) predicts that effort under prosocial incentive schemes should display scope insensitivity—namely, effort provided should be independent of incentive stakes.

Self-benefiting incentives, on the other hand, may not exhibit the same scope insensitivity. As demonstrated by Gneezy et al. (2011b), individuals under self-benefiting incentive schemes do respond positively to increases in incentive size once these incentives are already in place. Particularly, individuals already being compensated personally for their work will work harder when the payment stakes increase. The positive relationship between incentive size and effort under self-benefiting incentive schemes, in conjunction with the scope insensitivity under prosocial incentives, leads to our third hypothesis (Hypothesis 3): as the incentive size increases, individuals should both prefer and exert more or equal effort under self-benefiting rather than prosocial incentive schemes (provided the substitution effect exceeds the income effect).

We provide empirical support for these hypotheses using a novel experimental paradigm where real, costly effort is tied directly to either self-benefiting or prosocial incentives. We first look at whether individuals exert more effort under a prosocial incentive scheme than under a self-benefiting one when incentive stakes are low, and study whether this difference disappears or reverses when stakes are higher. We then examine if individuals anticipate these effects: whether they choose to work for charity more often than for themselves when incentives are low, and if this choice reverses when stakes are raised.

Our findings are consistent with our hypotheses: individuals work harder for charity than for themselves under low incentives or when no explicit incentives are provided. Effort provided under prosocial incentives, however, is no longer greater than under self-benefiting incentives when the stakes increase. This is primarily driven by the fact that individuals work significantly harder when the amount they are paid is greater, as economic reasoning would suggest, but are generally insensitive to the scope of the prosocial incentive. This insensitivity to scope is consistent with pure warm glow giving. Particularly, the effort they exert for charity does not change as the payment stakes increase. In addition, individuals' choice of incentive scheme is consistent with their choice of effort provision: they choose to work for charity when stakes are low and for themselves when the stakes increase. This contrasts with the inconsistency between choice and reported well-being found in previous work (Dunn et al., 2008).

2.2 Effort Experiment

We recruited 187 students from a university wide subject pool to participate in a 30 minute study. Participants were given a \$5 show up fee and could earn more depending on the condition and effort exerted.

Each session had between 6 to 8 participants who were randomly assigned to isolated computer terminals. There was at least one empty computer terminal between each participant to ensure privacy. A computer program randomized conditions within each session, and all instructions were displayed on each participant's individual computer screen.[1] Participants were asked to read the instructions to themselves and notify the experimenter if they had any questions. All questions were answered in private.

To measure effort, we asked subjects to squeeze a hand dynamometer that recorded force output in Newtons, twice. All participants were first asked to squeeze the device for 60 seconds to obtain a baseline measurement. The total force exerted was reported to the participants and acted as the baseline measure. Participants were then randomly assigned to one of five conditions in a 2 (For Self *vs.* For Others Incentives) $\times$ 2 (Low *vs.* High Incentives) between-subjects design, with an additional no incentive control condition.

After squeezing the hand dynamometer for 60 seconds to obtain a baseline measurement, participants were exposed to our manipulation. Participants were told that under the For Self incentive scheme (self-benefiting) they would receive the amount earned at the end of the experiment; the amount earned under the For Others incentive scheme (prosocial) would be donated to the Make-A-Wish Foundation (incentives are shown in Table 2.1). Each was then asked to squeeze the device again for 60 seconds after being matched into one of the 5 treatments: For Others and For Self incentive schemes were crossed with two incentive levels, Low and High. In the control condition, we simply asked participants to squeeze the device again for 60 seconds without giving them any explicit incentive to do so.

In the four incentive treatments, the amount earned by the participants or the charity was directly tied to effort. For example, participants in the For Self

[1] The experiment was programmed using the Qualtrics Research Suite.

treatment under High incentives earned \$2.00 cents for every 25k Newtons of force exerted during the treatment stage. They were not directly compensated for effort during the baseline stage. Earnings based on effort in the two incentive levels ranged between \$0.02 and \$0.35 in the Low treatments and \$2.73 and \$12.34 in the High incentive treatments.

As our dependent measure of effort, we used the ratio R of total force exerted during the treatment stage to that exerted in the baseline stage, which provides a normalized measure of effort that aims to control for individual characteristics such as gender and physical fitness.

Results

Table 2.1 lists averages of initial exerted effort for each of the 5 treatments. A preliminary test on initial effort revealed no between-group differences in baseline force exerted, $F(4, 182)=1.01$, *ns*. Similarly, pairwise comparisons show no significant differences in initial effort ($p > .1$ for all comparisons), suggesting that effort at the baseline stage did not differ by treatment.

Table 2.1 also lists the gender composition in each treatment; pairwise comparisons reveal no significant differences in compositions ($p > .2$ for all comparisons). In addition, there was no significant difference in the gender composition between the two High treatments and the two Low treatments plus Control ($t(185)=1.20$, $p = .23$). Running an OLS regression of the effort ratio R on a gender dummy (Male = 1; Female = 0) revealed no significant relationship between gender and our main dependent measure of effort ($\beta = .09$, $p = .497$). There were also no gender differences in the effort ratio R *within* any of the treatments ($p > .3$ for all comparisons).

We first examined the effect of the 4 incentive treatments (For Self *vs.* For Others Incentives) $\times$ (Low *vs.* High Incentives) on effort. We predicted that at Low incentive levels participants would exert more effort For Others than For Self (Hypothesis 1), with this difference disappearing or reversing as the incentive size increased (Hypothesis 3). Consistent with the first hypothesis, individuals indeed worked harder For Others (M=1.51, SD=.87) than For Self (M=1.14, SD=.34), $t(72)=2.37$, $p = .020$, when incentives were Low. However, in line with Hypothesis 3, the difference between incentive schemes disappeared when the incentive size was

High, $t(74)$=.96, p =.338. Indeed, a joint hypothesis test revealed that effort For Self under Low incentives was significantly lower than in the other three conditions, $F(3, 182)$=2.86, p=.038 (see Figure 2.1).

We further predicted that participants' effort would respond positively to increases in incentive size under the For Self incentive scheme, but be insensitive to the scope of incentives when working For Others (Hypothesis 2). In line with our second hypothesis and traditional economic reasoning, participants exerted significantly more effort under the For Self incentive scheme when incentives increased from Low to High (M=1.14, SD=.34 *vs.* M=1.74, SD=1.36), $t(70)$=2.58, p =.012. However, the increase in incentives did not change the level of effort exerted under the prosocial incentive scheme (M=1.51, SD=.87 *vs.* M=1.48, SD=1.03), $t(76)$ = -.13, p =.896.

To see whether the difference in difference between incentive schemes was significant as the size of incentives went up, we ran an OLS regression of the dependent variable R on the interaction between incentive scheme and incentive size. Our binary independent variables were the incentive scheme (For Others = 0; For Self = 1), incentive size (Low = 0; High = 1), and the interaction of the two. Neither incentive scheme (β = -.37, p =.107) nor incentive size (β = -.03, p =.898) had a significant influence on exerted effort. Importantly, however, we observed a significant interaction (β = .63, p =.049), suggesting that the relative effectiveness of prosocial incentive schemes is effected by the stakes involved.

Our control condition provided insight into the effectiveness of the incentive schemes relative to when no explicit incentives were provided. When no incentives were provided, individuals exerted less effort than under both prosocial incentive schemes, as well as under the self-benefiting incentive scheme when stakes were high (all ps < .02). However, when incentive levels were low in the self-benefiting incentive scheme, participants worked about as hard as when no incentives were provided (M=1.14, SD=.34 *vs.* M=1.07, SD=.29), $t(71)$=.90, p =.369. This offers further support for the notion that when incentives are low, prosocial incentive schemes may be superior to standard ones in motivating effort and performance.

To ensure that our results are robust to outliers, we conducted non-parametric permutation tests on the distributions of effort under the 4 incentive schemes. Using permutation methods, we constructed test statistics based on Schmid and Trede

(1996) and conducted one-sided tests for stochastic dominance and *separatedness* of the distributions (see also Ditraglia, 2006; Anderson et al., 2011). The test statistics identify the degree to which one distribution lies to the right of the other, taking into account both the consistency of differences between distributions (i.e. how often they crossed) and the magnitudes of the differences.[2]

The results of non-parametric distributional tests, p-values computed by Monte-Carlo methods with 10,000 repetitions, were consistent with the parametric tests above. We found a significant difference between the distributions of effort For Others and effort For Self under Low incentives (p =.027), as well as between the distribution of effort For Others and the control condition (p =.002). At High incentives, however, the distributions of effort For Others and For Self did not differ (p =.279), and both were significantly different than the distribution of effort in the control condition (both $ps < .01$).

To further rule out that our results were driven by outliers, we re-ran the analyses while winsorizing R at the upper bounds of the 95% confidence intervals of the treatments, which were 1.79, 1.80, 1.25 and 2.21 under the For Others (Low, High) and For Self (Low, High) incentive schemes, respectively.[3] The results were robust to winsorizing the data: individuals worked harder for charity than themselves at low incentive levels ($t(72)$=3.24, p =.001), and this difference disappeared when incentives were raised ($t(74)$=1.32, p =.19). Winsorizing R at a constant value (e.g., 3) did not change the results.[4]

2.3 Choice Experiment

We recruited 57 students from a university wide subject pool to participate in this study. Subjects were given a $5 show up fee and could earn more depending on the condition and effort exerted.

Effort was measured similarly to Study 1. All participants squeezed the hand

[2]For the data and a full description of how these tests were constructed, see Online Appendix (https://sites.google.com/site/alexoimas/prosocialincentives).

[3]The maximum Rs under the For Others (Low, High) and For Self (Low, High) incentive schemes were 4.55, 6.49, 1.90 and 5.88, respectively.

[4]See Online Appendix for further robustness checks

dynamometer for 60 seconds to obtain a baseline measurement and were informed of their initial force output. At the outset of the study, each participant was randomly assigned into one of two conditions: Low Incentives or High Incentives. They were then asked to choose one of two payment schemes, For Self (self-benefiting) or For Others (prosocial), prior to squeezing the device again for 60 seconds.

As in Study 1, effort exerted For Others benefited the Make-A-Wish Foundation and effort For Self was tied to payment received at the end of the experiment. Incentive under each condition are shown in Table 2.2.

Results

Averages of initial exerted effort for both treatments are listed in Table 2.2. A pairwise comparison revealed no significant difference in initial effort, $t(55)$=-1.00, *ns*. Gender composition did not differ by treatment, $t(55)$=-1.15, *ns*.

Our results suggest that individuals did indeed anticipate the benefits of prosocial vs. self-benefiting incentives as implied by Study 1 (see Figure 2.2). At Low incentive levels, 77% (23) of participants chose to work For Others, compared to 23% (7) who chose to work For Self. In contrast, at High incentive levels 15% (4) chose to work For Others, compared to 85% (23) who chose to work For Self. This difference was significant ($\chi^2(1) = 21.81$; $p <.001$).

Given the issues of selection in both incentive treatments, we cannot meaningfully compare the effort exerted by individuals under the two incentive schemes. Moreover, the resulting imbalance in cell sizes left these types of analyses severely underpowered. Nonetheless, results do appear to fit the general pattern outlined by our hypotheses, with the average R equaling 1.47 in the For Others condition versus 1.21 in the For Self when incentives were Low, and the same corresponding averages equaling 1.07 and 1.43 when incentives were High.

2.4 Discussion

Taking measured effort as a proxy for utility derived from the respective incentives, our findings suggest that individuals derive greater utility from prosocial spending than self-benefiting compensation, but only when the stakes are low. In addition, the utility derived is anticipated ex-ante: individuals *choose* to work for charity rather than themselves at low incentives, but prefer the opposite when stakes are raised.

Our results are consistent with a model of pure warm glow giving and help inform the literature on optimal incentive design. Traditional economic reasoning suggests that direct monetary payment will provide the strongest incentives. However, these self-benefiting monetary incentives have been shown to backfire, particularly when the payment stakes are low (Deci, 1971; Gneezy and Rustichini, 2000): monetary incentives may decrease prosocial behavior (Ariely et al., 2009), reduce socially efficient contributions to a public good (Fuster and Meier, 2009), and result in lower effort provision than could be achieved with no payment scheme at all (Gneezy et al., 2011b; Heyman and Ariely, 2004). These findings suggest that there may be situations where direct monetary compensation is not the optimal incentive scheme to motivate individuals to exert effort.

Suppose that a company offers incentives to its employees to lose weight. Providing low direct compensation, e.g. $5 per pound lost, may not be enough to motivate individuals to adopt a healthier routine, and may in fact crowd out positive steps they were already taking. Indeed, recent evidence on the effectiveness of paying people directly for weight loss has been mixed at best (Kullgren et al., 2013). Our results suggest that prosocial incentives may provide stronger motivation for individuals to lose weight than "standard," self-benefiting incentives. Particularly, if the budget for a certain task or project is small, organizations should take advantage of the scope insensitivity of prosocial incentives by tying small charitable contributions to performance rather than compensating individuals directly.

Future research should explore whether making the contributions under a prosocial incentive scheme public or private has an effect on effort provision. Given the social signaling value of prosocial behavior, a prosocial incentive scheme may be even more effective in motivating individuals when effort-contingent contributions are public. In addition, since many employment contexts carry the expectation of direct compensation, it would be useful to compare the performance of individuals under a prosocial incentive scheme to when prosocial and self-benefiting incentives are mixed (Ariely et al., 2009; Tonin and Vlassopoulos, 2012, 2010).

2.5 Figures

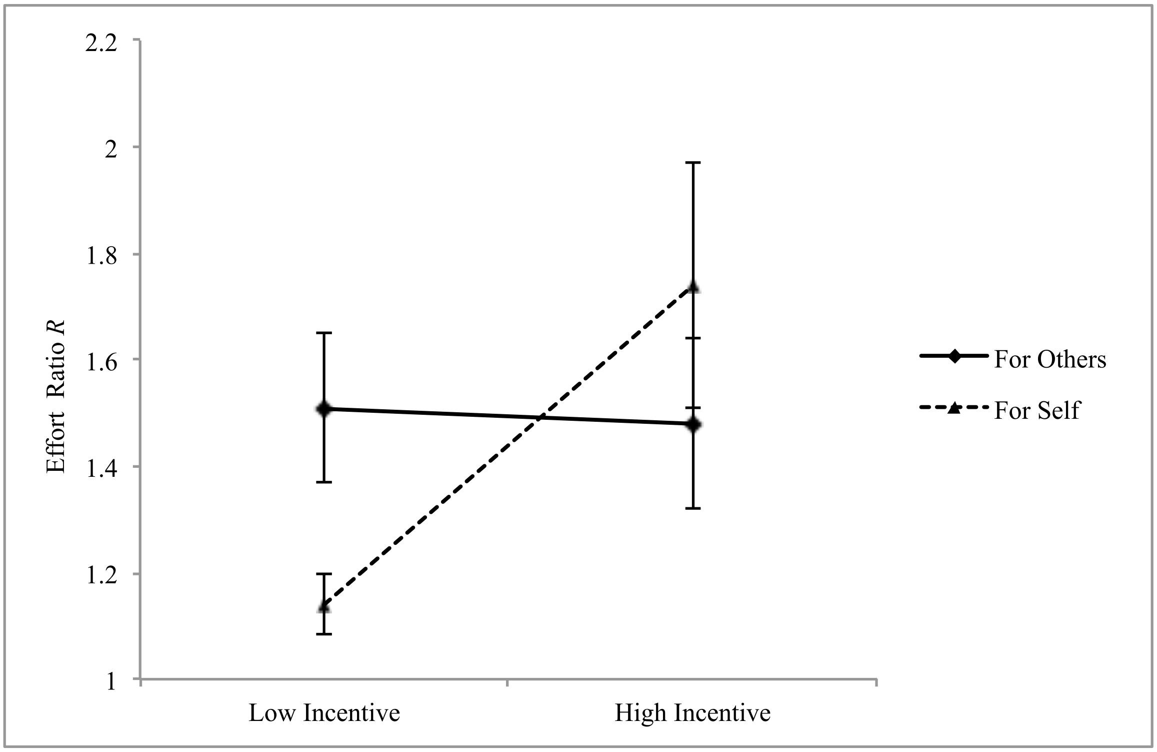


Figure 2.1: Effort Ratio R by Treatment

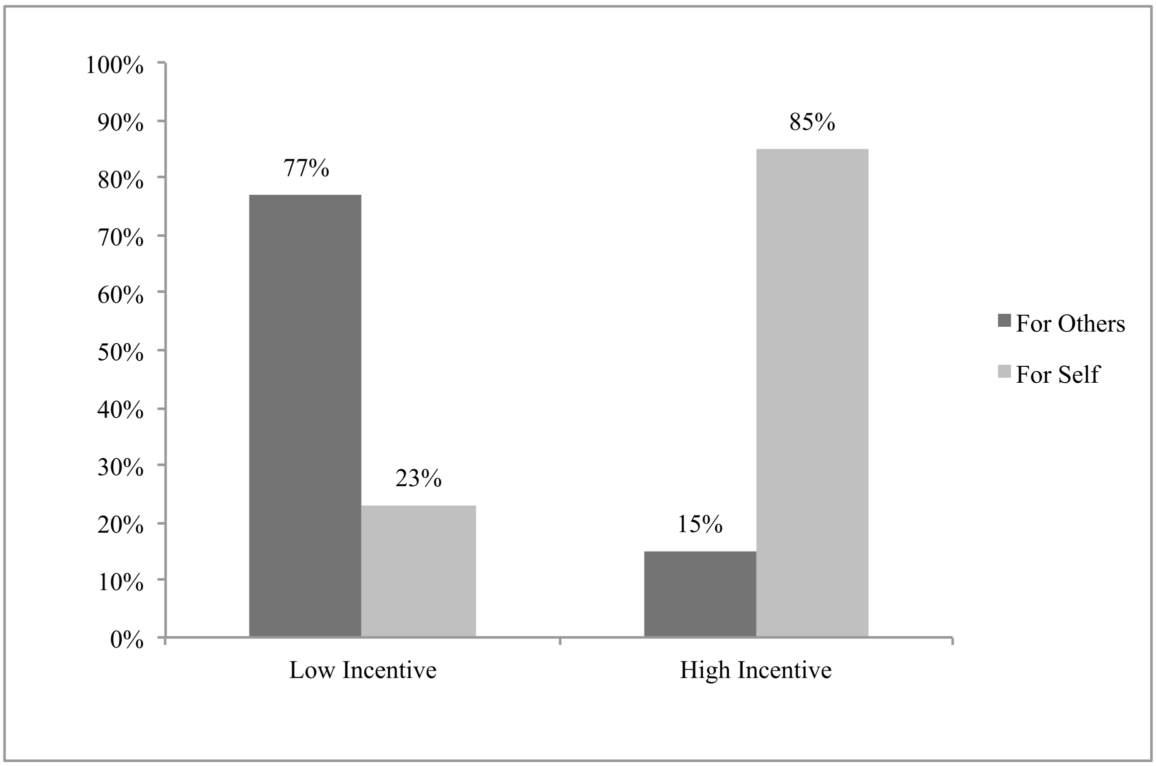


Figure 2.2: Percentage of Participants Choosing Incentive Scheme

2.6 Tables

Table 2.1: Treatment Summary

Treatment	**Incentive** (*per 25k Newtons*)	$\boldsymbol{N}$	**Gender** (*% Male*)	**Initial** (*Newtons*)	$\boldsymbol{R}$
For Others *Low*	$0.05	38	39%	65,620	1.51
For Others *High*	$2.00	40	55%	57,991	1.48
For Self *Low*	$0.05	36	44%	68,450	1.14
For Self *High*	$2.00	36	53%	59,598	1.74
Control	-	37	51%	64,836	1.07

Table 2.2: Treatment Summary

Treatment	**Incentive** (*per 25k Newtons*)	$\boldsymbol{N}$	**Gender** (*% Male*)	**Initial** (*Newtons*)	$\boldsymbol{R}$
Low Incentives	$0.05	30	37%	56,621	1.41
High Incentives	$2.00	27	52%	64,387	1.37

Chapter 3

The Materazzi Effect and the Strategic Use of Anger in Competitive Interactions

3.1 Introduction

Zinedine Zidane is considered one of the greatest soccer players of all time, leading the French team to victory in the 1998 World Cup and the 2000 European Championship. In the 2006 World Cup in Berlin, the French team did well under Zidane's leadership and reached the finals, where it played against the Italian team. This game was to be the last of Zidane's career before his retirement from soccer.

The score was 1:1 and the game went into overtime. With 10 minutes left, Marco Materazzi, an Italian defender, pulled Zidane's shirt. Zidane responded with, "If you want my shirt that badly, I'll give it to you at the end of the match"—nothing out of the norm, the usual trash talking between rival team members. Then Materazzi shot back: "I'd prefer your whore of a sister." Zidane lost it and headbutted Materazzi in the chest. This move was Zidane's last on-field act as a soccer player.

The game went on to penalty kicks without Zidane—the French team's best kicker—and the Italian team won. Both Zidane and Materazzi will likely be remembered more for this headbutt than for anything else either did in his career. Instead of leaving in glory as a second-time World Cup champion, Zidane became a parody, and Materazzi, a hero.

What happened to Zidane during those few seconds? And did Materazzi anger him strategically, expecting the strong reaction? What is clear is that the Italian team's chance of winning the game increased significantly as a result of Materazzi's insult.

In this paper we explore the strategic use of emotions, particularly anger, in interactive situations. We outline and test predictions regarding when the use of anger benefits the offender and when it backfires. Our main argument is that angering others affects their behavior, either helping or hurting performance; individuals understand these effects and employ anger strategically in interactions.

Recent work has shown that emotions have a substantial effect on economic behavior and decision making (Bracha and Brown, 2012; Koszegi, 2006; Loewenstein, 2000, 1987; Neilson, 2009; Gino and Schweitzer, 2008). For example, individuals are kinder to others so as to avoid feelings of guilt (Dufwenberg and Gneezy, 2000; Battigalli and Dufwenberg, 2009; Charness and Dufwenberg, 2006) and make more conservative investment choices when anxious or afraid (Loewenstein et al., 2001; van Winden et al., 2011).

Feelings of anger have been shown to be important factors in social punishment (Fehr and Gächter, 2002; Fuster and Meier, 2009; Hopfensitz and Reuben, 2009; Sutter et al., 2010). Angered individuals are more willing to reject offers in an ultimatum bargaining game and consequently make less money (Andrade and Ariely, 2009; Pillutla and Murnighan, 1996). Anger also significantly effects behavior in the repeated prisoner's dilemma games (Dreber et al., 2008; Fudenberg et al., 2012) and power-to-take games (Bosman et al., 2005). In addition, people are more willing to become angry when they expect to face situations that require aggressive behavior (Tamir et al., 2008).

Importantly, intentionality of the offense seems to be necessary for anger to be induced (Offerman, 2002). Individuals are significantly more likely to punish an intentionally hurtful choice than if the same choice was made unintentionally. The difference in negative reciprocation is primarily driven by an increase in anger when the hurtful act was viewed as intentional (Blount, 1995; Charness, 2004).

Being able to express anger through other channels appears to mitigate the need to punish unfair treatment, with individuals significantly less likely to punish when they are able to communicate their feelings to the offender (Xiao and Houser, 2005). In addition, strategic use of the *appearance* of anger seems to have substantial influence on the behavior of others, increasing the other party's initial offers in ultimatum bargaining games and concessions in negotiations (Andrade and Ho, 2009; van Kleef et al., 2004; Gibson et al., 2009).

Little work, however, has been done to explore how people in strategic interactions exploit the behavioral impact of *experienced* emotions such as anger. To test whether individuals anger others strategically, we designed an experiment where two players are matched and compete in one of two games. Prior to the competition, one is randomly chosen to have the option to anger his opponent. One of the games is strength-based, where we expected angered individuals to be more successful; the other game is more cognitively demanding, where anger is expected to have a negative effect on performance. In turn, we hypothesized that in the mental game angering one's opponent benefits the offender while in the strength game it may backfire.

To identify the effect of anger and rule out alternate explanations, we also introduced treatments with a delay between the potentially angering action and the subsequent competition. Given prior evidence on the temporal dynamics of emotions (Loewenstein, 1996), we expected the behavioral effects of anger to diminish after individuals had a chance to "cool off." Particularly, we hypothesized that angered and non-angered individuals would behave similarly in both games after the delay. Further, if individuals anticipate the mitigating effects of delay, the strategic advantage of angering one's opponent differentially between the two games should also be weakened.

3.2 Experiment

3.2.1 A Strength Game

In the strength game participants compete on the total force they could produce over time by squeezing a hand dynamometer that measured handgrip strength in Newtons. Research has shown that anger positively affects performance on such strength-based tasks (Davis et al., 2010).

We recruited male undergraduates (N=140) from a university-wide subject pool to participate in this experiment. Randomly paired participants compete twice in a game of strength by gripping the dynamometer for two one-minute rounds. They were told that one of the two rounds would be randomly chosen and the individual who exerted the most total force would win $5, and the other would receive nothing. Participants did not learn of the outcome until the end of the experiment.

After the first round, one individual in each pair was told that our behavioral lab had a number of boring administrative tasks to complete and that he was given the opportunity to choose how long his opponent would stay after the experiment to work on these tasks. We termed the chosen individual the Decision Maker (DM) and his opponent the Worker. The DM could choose to make the Worker stay between 0 to 20 minutes, in 5-minute intervals. The DM received an additional, relatively small amount of $0.50 for every 5 minutes he assigned to the Worker (maximum of $2) while the Worker received no additional compensation.

We informed the DM that the Worker would know the choice and incentives he faced, as well as the decision he made, before beginning the second round of the strength game. Note the maximum earnings for making Worker's stay was $2, compared to the prospective earnings of $5 from winning the subsequent competition.

We also tested a delay condition, in which the DM and Worker were both told that after the DM's choice was revealed there would be 10-minute break where each would complete an unrelated task before beginning the competition.

Once the Worker was informed of the DM's choice, they competed in the strength game either without a break (No Delay), or after a 10-minute filler task of neutral anagrams (Delay).

The DM's choice served as our anger manipulation. We ran a test which confirmed that Workers who had to stay after the experiment for longer amounts of time were angrier than those who had to stay for less time or not at all, and that DMs anticipated this effect. See Appendix for details and analysis.

We predicted that those who were angered between rounds—assigned to stay for an extended period of time after the experiment—would perform better in the strength game than those not angered.

Results

For our analysis we used the ratio R between the total force exerted in rounds 2 and 1 as the dependent measure. Examining ratios allows us to minimize the effects of individual characteristics such as physical fitness.

In line with our prediction, Workers in the No Delay treatment who were assigned the full 20 minutes (M=1.24) performed better than Workers assigned less time (M=.97; $t(38)$=2.91, p=.006). Workers assigned 10 minutes (M=.98) behaved similarly to those assigned 0 minutes (M=.96; $t(19)$=.11, p=.92). In turn, the 10-minute group performed worse than those assigned the full 20 minutes ($t(29)$=2.25, p=.032).

As predicted, in the Delay treatment Workers performed similarly irrespective of whether 20 minutes or less than 20 minutes were assigned (M=.93 *vs.* M=.91, respectively; $t(28)$=.16, p=.87). Further, those assigned the full 20 minutes in the Delay treatment performed significantly worse than those assigned 20 minutes in the No Delay treatment ($t(29)$=2.82, p=.009).

Given these findings, angering one's opponent in the strength game with No Delay would not be strategically smart since angered participants performed better than those who were not. Using our data, we can examine the chances of winning in expectation when facing the median DM. Our results indicate that assigning 20 minutes to the Worker in the No Delay treatment cost the DM $1.45 in expectation relative to assigning less minutes—a significant difference ($t(38)$=2.17, p=.036). In contrast, assigning 20 minutes in the Delay treatment did not cost the DM any more in expectation than assigning less minutes (-$0.34; $t(28)$=.36, p=.72). Figure 3.1 shows expected returns of the DM for each condition relative to the expected return if he had the same probability of winning or losing, $2.50. See Appendix for full analysis.

3.2.2 A Duel Game

The second game was designed to be more cognitively demanding, where optimal performance required computation and patience. Given evidence on the detrimental effect of anger on depth of processing and self-restraint, we predicted anger would impair performance in this game (Knapp and Clark, 1991; Tiedens and Linton, 2001).

Randomly paired participants played a computerized duel game (N=120). Before the game was played, a set of instructions was given to each participant and read aloud. The same anger manipulation was used as in the strength game, with one player chosen at random to be the DM and his opponent the Worker. Participants were matched into either the Delay or No Delay treatment. After the DM made his choice and the Worker learned of it, the two competed in the duel game.

In a computerized game, participants started 20 steps apart with one bullet, meaning each had one chance to shoot his opponent. Period 1 started with player 1 making the choice to either "step forward" or "shoot" while player 2 waited for him to decide. If player 1 chose "step forward," Period 2 would begin and the distance between the players would decrease by one step (from 20 to 19). Player 2 would then face the same choice, and so on until one of the players chose to shoot. The players were told the probability of hitting their opponent in every Period. This probability increased monotonically as the distance between the two players decreased. If players were 20 steps apart, the probability was 0. The probability went up by .03125 until Period 16, when it was .5, and then went up by .125 until Period 20, when it was 1.

If a player shot and hit his opponent, he would get $5 and his opponent $0; if he missed, he would get $0 and his opponent $5. A participant won the game if he shot and successfully hit his opponent or if his opponent shot and missed him.

Since there are only two outcomes, win or lose, and conditional on the payoffs, each player had an optimal strategy that was independent of risk attitudes: "step forward" until the probability reached 0.5 and then "shoot." Particularly, given that players made decisions in alternating periods, the Worker's optimal strategy was to "step forward" through Period 15 and "shoot" in Period 17 (see Appendix for more details).

Given the expected effect of anger on depth of reasoning and patience, we predicted that angered participants would behave more sub-optimally than those not angered: they would be more likely to shoot first (and too early).

Results

If the Worker shot first, he did so significantly earlier than Period 17 in both the No Delay (M=13.11; $t(16)$=4.78, $p<.001$) and Delay (M=13.4; $t(9)$=2.86, p=.02) treatments. There was no significant difference between the Delay and No Delay treatment in what round the shot was fired.

In line with our prediction, angered individuals in the No Delay treatment were more likely to shoot first and too early than those who were not angered. Workers assigned the full 20 minutes shot first 70% of the time, while those assigned less time shot first 25% of the time ($t(30)$=-2.66, p=.01).

In the Delay treatment, there was no significant difference in the tendency to shoot first based on whether 20 minutes or less than 20 minutes were assigned (38% *vs.* 33%, respectively; $t(26)$=.22, p=.83). Additionally, those assigned 20 minutes without delay were more likely to shoot first than those assigned 20 minutes with delay ($t(32)$=2.11, p=.04).

Given that angered individuals were more likely to make a sub-optimal choice, it follows that angering one's opponent in the duel game with No Delay was strategically smart. This finding is in contrast to the strength game where angered individuals performed better. In fact, assuming the DM follows his optimal strategy, by assigning 20 minutes to the median Worker in the No Delay treatment he stood to gain \$0.28 in expectation relative to assigning less minutes ($t(30)$=2.66, p=.01). In contrast, assigning 20 minutes in the Delay treatment did not change expected payoffs relative to assigning less (\$0.13; $t(26)$=1.32, p=.20). Figure 3.2 shows expected returns of the DM for each condition relative to the expected return if he had the same probability of winning or losing, \$2.50. See Appendix for full analysis.

3.2.3 Anger as a Strategy

Our results suggest anger that affected performance differently depending on the game, impairing performance in the duel game and enhancing it in the strength game. These behavioral effects were mitigated, however, when individuals had an opportunity to "cool off" before competing.

Using anger strategically implies that individuals should be more willing to anger opponents when it increased their chances of winning. Particularly, DMs in the duel game with No Delay should assign more minutes to the Workers than in the strength game. Our results are consistent with this prediction. In the strength game, 45% of Workers were assigned the full 20 minutes, whereas in the duel game 63% of Workers were assigned the full time. In contrast, 20% of Workers in the strength game were assigned no time, compared to 6% of Workers in the duel game.

A regression analysis confirms the differential tendency to anger between games. DMs playing the duel game assigned Workers significantly more minutes than those playing the strength game (β=3.41, p=.045). These results suggest that participants anticipated the behavioral effects of anger and used it strategically—they were more willing to anger their opponents when it had a detrimental effect on performance.

Individuals also seemed to anticipate the mitigating effect of delay on the anger's behavioral effects. Particularly, in the treatments where the DM was told there would be a break between the Worker learning of the minutes assigned and the competition, the type of game no longer had a significant effect on the number of minutes assigned (β=.29, p=.46).

3.3 Conclusion

Our findings suggest that anger has discernible behavioral effects that depend on the context: anger had a positive effect on performance when the game involved strength and a negative effect when the game was more mental. Individuals seemed to understand these effects, choosing to anger opponents more when competing in the latter than the former. This finding suggests that participants offended strategically.

Although we focus primarily on anger, one can apply our findings to other emotions as well. For example, recent work has demonstrated the positive effects of guilt on charitable behavior (Gneezy et al., 2014). Organizations looking to maximize donation revenues can exploit these emotional effects by inducing guilt strategically in potential donors. Additionally, given the significant effects of emotions on behavior, our findings suggest that incorporating emotions such as anger into models of strategic interactions would provide better understanding of such environments.

Chapter 3, in part, has been submitted for publication of the material as it may appear in the Proceedings of the National Academy of Sciences, 2014, Gneezy, Uri; Imas, Alex O., National Academy of Sciences, 2014. The dissertation author was the co-primary investigator and author of this paper.

3.4 Appendix

3.4.1 Anger Manipulation

To ensure the validity of our anger manipulation, we ran a test which confirmed that Workers who had to stay after the experiment for longer amounts of time were angrier than those who had to stay for less time or not at all, and that DMs anticipated this effect.

A separate group of participants (N=70) were recruited and matched into pairs. In each pair, one was assigned the role of DM and the other was assigned the role of Worker. We used the strategy method to elicit the Worker's anger response to being assigned a particular amount of minutes to work, and the DM's expectations for how angry the Worker would be after seeing their choice of minutes assigned.

Each Worker was told that the DM would assign a number of minutes for them to work and this choice would actually be carried out. Workers were then asked how angry they would feel on a scale of 1-7 if they were assigned to work a particular number of minutes. Running a regression of reported anger on the DM's choice of minutes assigned (in 5-minute blocks) with standard errors clustered at the individual level revealed a significant relationship, (β=.81, p<.001): the more time Workers were assigned to work, the angrier they reported feeling.

Similarly, DMs were asked how angry they expected the Worker to be, on a 1-7 scale, after being assigned a particular number of minutes to work and then made a choice that would actually be carried out. A regression of minutes assigned (in 5-minute blocks) on expected anger with standard errors clustered at the individual level showed that DMs expected Workers to be angrier the more minutes were assigned (β=1.11, p<.001).

3.4.2 Strength Game

Minutes Assigned and Performance

Categorical Analysis

For our analysis we used the ratio R between the total force exerted in rounds 2 and 1 as the dependent measure. Examining ratios allows us to minimize the effects of individual characteristics such as physical fitness.

We regressed the ratio of effort R of the Workers on the number of minutes assigned. In the No Delay condition, Workers exerted significantly more effort the more minutes they were assigned (β=.015, p=.018). In the Delay condition, there was no significant relationship between effort and number of minutes assigned (β=.004, p=.402).

Binary Analysis

Workers in the No Delay treatment who were assigned the full 20 minutes (M=1.24) performed better than Workers assigned less time (M=.97; t(38)=2.91, p=.006). Workers assigned 10 minutes (M=.98) behaved similarly to those assigned 0 minutes (M=.96; t(19)=.11, p=.92). In turn, the 10-minute group performed worse than those assigned the full 20 minutes (t(29)=2.25, p=.032).

In the Delay treatment Workers performed similarly irrespective of whether 20 minutes or less than 20 minutes were assigned (M=.93 *vs.* M=.91, respectively; t(28)=.16, p=.87). Further, those assigned the full 20 minutes in the Delay treatment performed significantly worse than those assigned 20 minutes in the No Delay treatment (t(29)=2.82, p=.009).

3.4.3 Expected Return

To calculate the expected gains or loses from a particular choice of minutes assigned, we took the ratio of force output R of the average DM and compared his chances of winning against a Worker who was assigned either 20 minutes or assigned less than 20 minutes. We did this for both the No Delay and Delay treatments. The average R of a DM in the No Delay treatments was .901 and the average R of a DM in the Delay treatment was .897. Table 1 summarizes the probabilities of having an R greater than a Worker conditional on the minutes assigned and the delay treatment.

These probabilities are multiplied by the prize for winning ($5) to calculate the expected value of the competition for the DM. Figure 3.3 summarizes the expected return of assigning a particular number of minutes in each of the two delay conditions.

3.4.4 Duel Game

Optimal Strategy

The optimal strategy for the duel game can be calculated using backward induction. In period 20 there is 1 step between the two players. The probability of hitting the opponent conditional on choosing "Shoot" is .875, so the expected value (EV) of shooting in period 20 is $5×.875=$4.38. The game automatically ends after period 20; if neither player chooses to shoot, one player is randomly selected to receive $5. Hence, the EV of stepping forward in period 20 is $2.50. Since the EV of shooting is higher than the EV of stepping forward, the optimal action in period 20 is to choose "Shoot." In period 19, the player knows that in period 20 his opponent will choose "Shoot." His EV of choosing "Step forward" is then the probability that his opponent will shoot and miss in the next period (.125) multiplied by the prize ($5×.125=$.63). His EV of choosing "Shoot" is the probability of a successful shot (.75) multiplied by the prize ($5×.75=$3.75). Since the EV of shooting is greater than the EV of stepping forward, the optimal action in period 19 is "Shoot." The result of iterating this calculation backwards for every period is that the EV of "Shoot" is greater than the EV of "Step forward" for periods 17-20, the EVs are the same in period 16, and the EV of "Step forward" is higher than the EV of "Shoot" for periods 1-15 (see Table 2 and Figure 3.4). Since the Worker makes the choice in odd periods, his optimal strategy is to "Step forward" periods 1-15 and "Shoot" in periods 17-19.

Minutes Assigned and Performance

If the Worker shot first, he did so significantly earlier than Period 17 in both the No Delay (M=13.11; $t(16)=4.78$, $p<.001$) and Delay (M=13.4; $t(9)=2.86$, $p=.02$) treatments. There was no significant difference between the Delay and No Delay treatment in what round the shot was fired.

Categorical Analysis

We regressed a dummy variable corresponding to whether the Worker shot first or not (1=Worker shot first; 0=DM shot first), on the number of minutes assigned. In the No Delay condition, Workers were more likely to shoot first – prior to the optimal period – the more minutes they were assigned (β=.036, p=.013). In the Delay condition, there was no significant relationship between the propensity to shoot first and number of minutes assigned (β=.006, p=.527).

Binary Analysis

In the No Delay treatment, Workers assigned the full 20 minutes shot first 70% of the time, while those assigned less time shot first 25% of the time ($t(30)$=-2.66, p=.01). In the Delay treatment, there was no significant difference in the tendency to shoot first based on whether 20 minutes or less than 20 minutes were assigned (38% *vs.* 33%, respectively; $t(26)$=.22, p=.83). Additionally, those assigned 20 minutes without delay were more likely to shoot first than those assigned 20 minutes with delay ($t(32)$=2.11, p=.04).

Expected Return

To calculate the expected gains or loses from a particular choice of minutes assigned, we assume that the DM is playing the optimal strategy against the average Worker conditional on the minutes assigned and delay treatment. If the Worker is not observed to shoot first, we assume he is playing the optimal strategy. In the No Delay treatment, if he was the first to shoot, the average Worker shot in period 13.14 when assigned 20 minutes and in period 13 if assigned less than 20 minutes; Workers assigned 20 minutes shot first 70% of the time and while those assigned less than 20 minutes shot first 25% of the time ($t(30)$=-2.66, p=.01). Using the expected values listed in Table 2, a DM in the No Delay treatment assigning 20 minutes to the Worker stood to gain \$0.43 while a DM assigning less time stood to gain \$0.16 in expectation ($t(30)$=2.66, p=.01). In the Delay treatment, the average Worker who shot first did so in period 12.67 when assigned 20 minutes, and in period 14.5 when assigned less; if assigned 20 minutes, Workers shot first 38% of the time, if assigned less than 20 minutes, Workers shot first 33% of the time ($t(26)$=.22, p=.83). Using the expected values listed in Table 2, in the Delay treatment a DM assigning 20 minutes stood to gain \$0.23 and \$0.10 if he assigned less ($t(26)$=1.32, p=.20). Figure 3.5 summarizes the expected return of assigning a particular number of minutes in each of the two delay conditions.

3.4.5 Between Game Comparisons

Number of Minutes Assigned by Game

See Figure 3.6.

Discussion of Differences Between Games

Individuals completed the strength game twice in order to obtain a baseline level of strength in the first round of the competition. By dividing the second round output by the first round output, we created a normalized measure of performance which controlled for individual characteristics such as baseline strength. This was done to improve the power of the analysis. Importantly, individuals in the Strength game were not told their relative performance until the end of the experiment, and hence Workers had the same information on relative performance after receiving the DMs choice as those in the Duel game. In addition, since individuals were only compensated for one round of the competition, incentives were the same in the Duel and Strength game. We did not have a first round of the Duel game because a normalized measure was not necessary and we were concerned that a first round choice to shoot could act as an anchor for the second round choice.

In addition, the Delay conditions served as controls for any differential characteristics of the games apart from the strategic benefits of anger. Particularly, individuals played two rounds of the Strength game in both the Delay and No Delay conditions; similarly, they played one round of the Duel game in both Delay and No Delay conditions. Our results suggest that anger was used strategically and effectively only in the No Delay conditions.

3.5 Figures

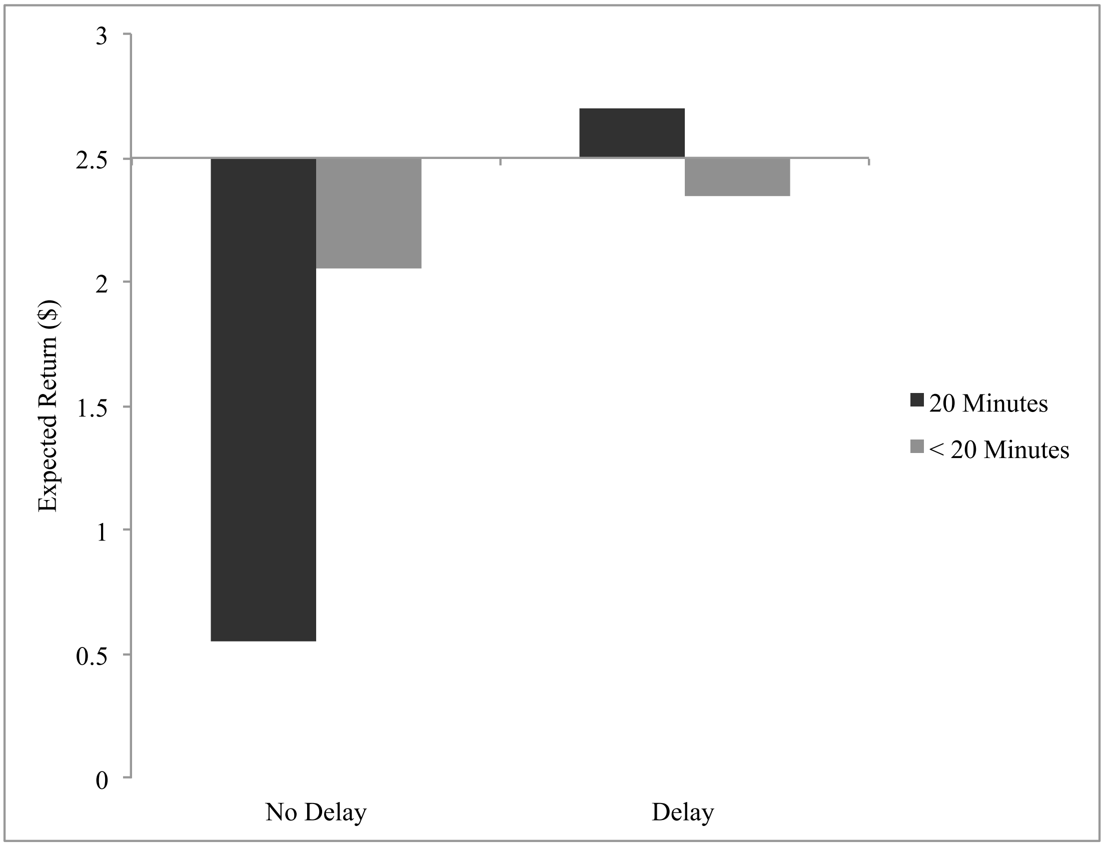


Figure 3.1: Decision Maker's Expected Return in Strength Game

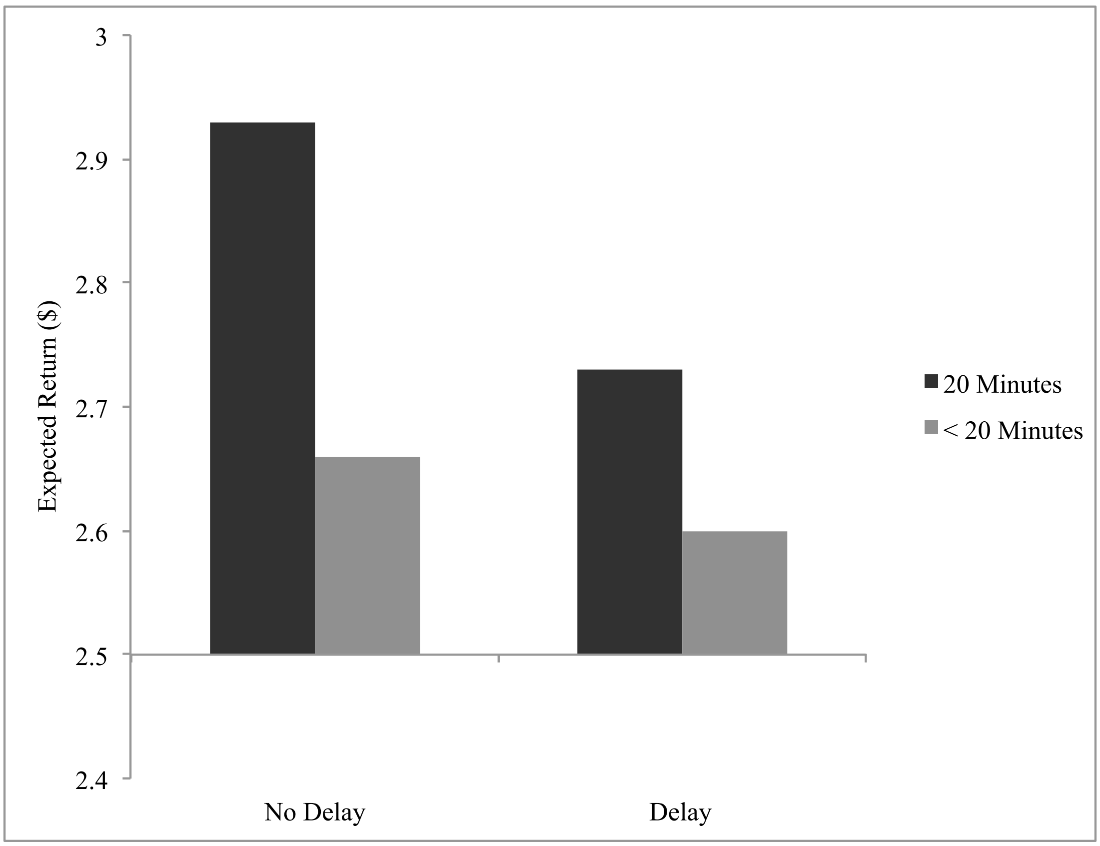


Figure 3.2: Decision Maker's Expected Return in Duel Game

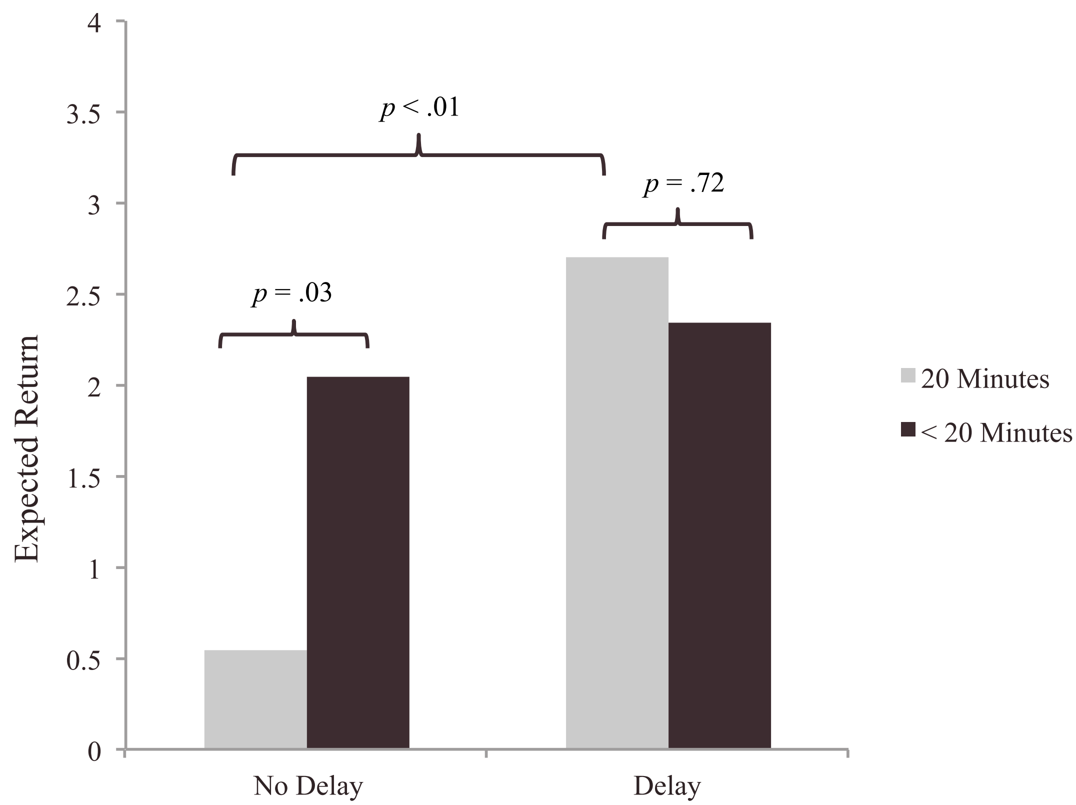


Figure 3.3: Expected Return

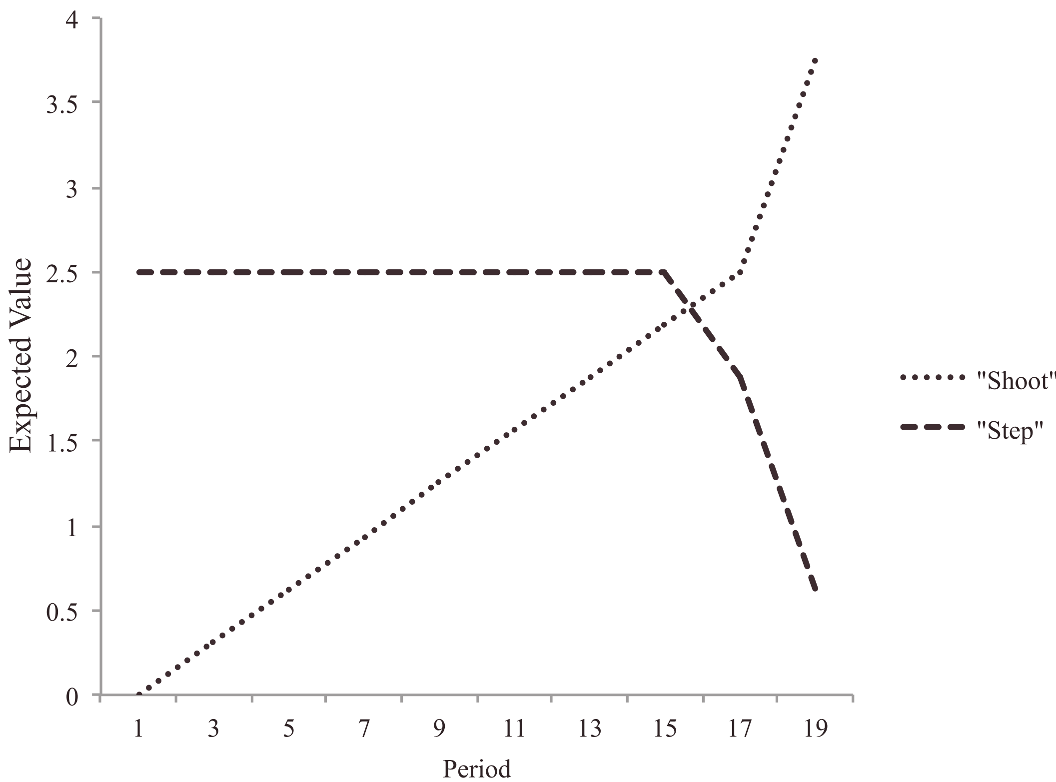


Figure 3.4: Expected Value

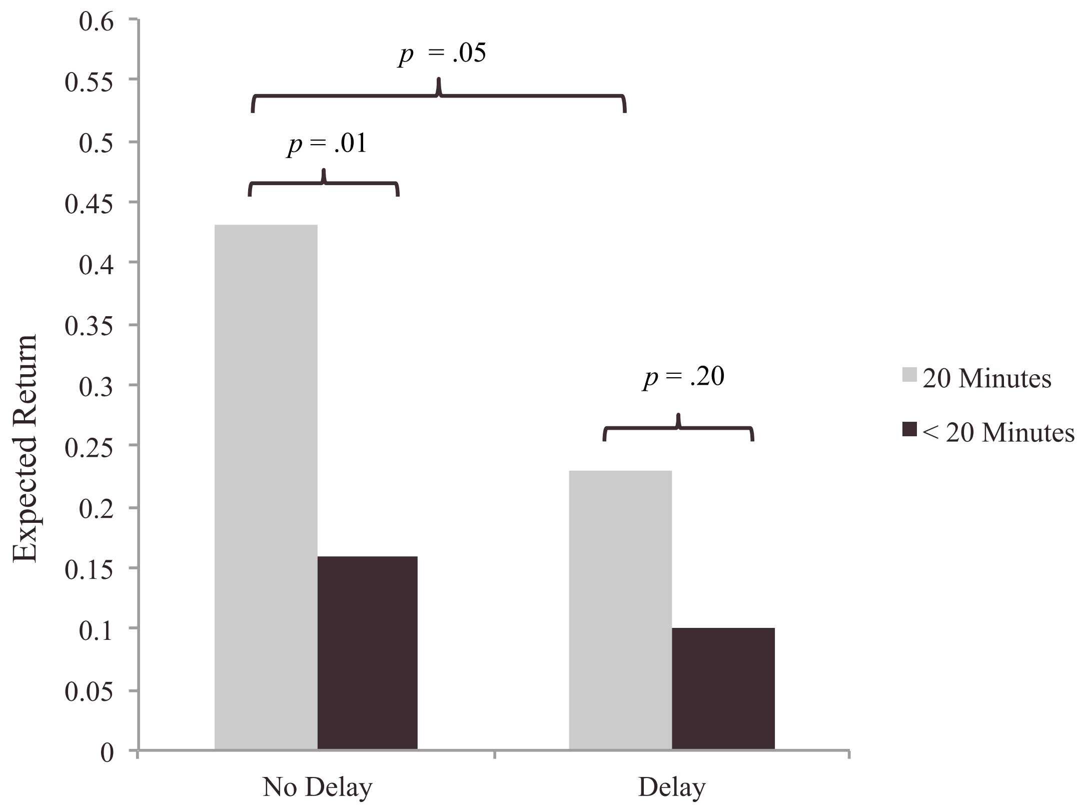


Figure 3.5: Expected Return

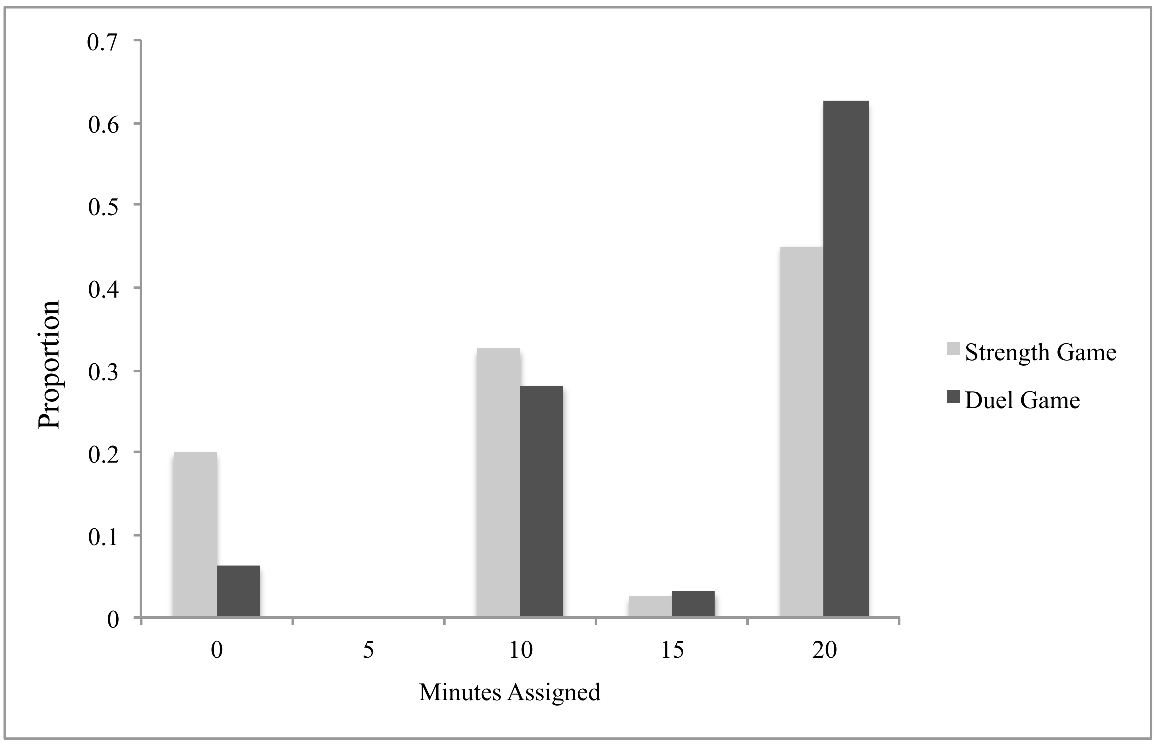


Figure 3.6: Number of Minutes Assigned by Game

3.6 Tables

Table 3.1: Probability of Winning for Median DM

	20 Minutes	**<20 Minutes**
No Delay	.11	.41
Delay	.54	.47

Table 3.2: Probability of Winning for Median DM

Period	**EV "Shoot"**	**EV "Step"**	**Net EV**	**Optimal Choice**
1	0	2.5	-2.5	Step
3	.3125	2.5	-2.19	Step
5	.625	2.5	-1.88	Step
7	.938	2.5	-1.56	Step
9	1.25	2.5	-1.25	Step
11	1.56	2.5	-.94	Step
13	1.88	2.5	-.625	Step
15	2.19	2.5	-.313	Step
17	2.5	1.88	+.625	Shoot
19	3.75	.63	+3.125	Shoot

Chapter 4

Conscience Accounting: Emotion Dynamics and Social Behavior

> Terrible is the Temptation to do Good – Bertolt Brecht, *The Caucasian Chalk Circle* (1944)

4.1 Introduction

This paper argues that dynamic fluctuations of guilt shape people's prosocial motivations. We present evidence that people who first made an unethical choice were then more likely to donate to charity than those who did not. We interpret these results in the context of our theory, which offers a new explanation for why people donate to charity: to offset a feeling of guilt associated with recent "bad" actions. We term this result "conscience accounting."

Throughout history, institutions have been built to take advantage of the effects of guilt on charitable behavior and to enable individuals to account for their conscience. For example, around the time of the Second Temple—500 B.C. to 70 A.D.—Jewish leaders formalized the use of chatot (sins) and ashamot (guilt) offerings as atonement for transgressions. The medieval Catholic Church adopted a similar technique when it began to grant "indulgences" that absolved an individual of sins through a system of "tariff penances." Today, Mass in the Catholic Church typically involves congregants reciting a prayer called the Confiteor (Mea Culpa) in which they confess, and are in turn reminded of, their sins. Afterwards, the Church solicits alms by requiring congregants to pass around a collection plate.

These kinds of institutions appear to take advantage of individuals' self-imposed moral constraints and the need to account for past transgressions through compensatory behavior. However, by providing people with the explicit opportunity to relieve their conscience, such practices may actually increase unethical behavior by lowering its cost. In this paper we present evidence for both effects, showing that individuals are more likely to be charitable within a temporal bracket after an unethical choice and that the knowledge of subsequent prosocial opportunities increases unethical behavior ex-ante.

Our paper begins with the observation that an individual experiences aversive feelings of guilt after she violates an internalized norm or acts in a way that she views as unethical.[1] After the initial increase, such guilt depreciates over time as the individual's emotional state gradually reverts back to the original "cold" one (Elster, 1998). The identification of our proposed hypotheses relies on the fact that such emotional fluctuations lead to a time-inconsistency in the individual's ex-post prosocial behavior. A moral transgression creates a temporal bracket where the immediate onset of guilt creates a greater propensity for prosocial behavior which diminishes over time. We term this emotional response conscience accounting, and this systematic change in social preferences also helps us identify norm violations in choice behavior.

Being at least partially aware of their emotional responses, our mechanism implies that individuals may adjust their ex-ante behavior in accordance to available future prosocial opportunities. Since donating after a norm violation will make the guilty person feel better, if the charitable opportunity is small or limited, then knowing that this opportunity exists may *encourage* unethical behavior in the present. However, if the charity option is large or unlimited, individuals may initially choose to avoid or delay charitable opportunities after moral transgressions because they fear that their guilty self will be overly generous.

[1] Guilt is associated with moral transgressions (Baumeister et al., 1994), and the desire to avoid violating the expectations of others has been established as equilibrium behavior within a game theoretic framework (Battigalli and Dufwenberg, 2007, 2009; Charness and Dufwenberg, 2006; Dufwenberg and Gneezy, 2000). Guilt is also considered an aversive feeling that discourages norm violations (Akerlof and Kranton, 2000), and once experienced, facilitates social behavior (Amodio et al., 2007; de Hooge et al., 2007).

We test our behavioral hypotheses using two experimental paradigms, finding support for the prediction of conscience accounting: individuals who achieved a given payoff by deceiving or stealing were more likely to donate to charity than those who achieved the same payoffs in a more ethical manner. In addition, our results show that this effect occurs within a temporal bracket, such that the increase in prosocial behavior is greatest directly after the unethical act and decreases with time. We also find that individuals anticipate these effects, lying more when they know that a limited donation opportunity will follow.

Our results have a direct application to charitable contributions and volunteering behavior, suggesting an additional explanation for why people donate their money and time. The willingness to give has puzzled economists for decades; not only because it contradicts the assumption that people are fueled solely by self-interest, but also because it does not seem to be driven by one simple alternative (Andreoni, 1990; Becker, 1976; Cappelen et al., 2013; Meier, 2007; Ariely et al., 2009; Small and Loewenstein, 2003; Vesterlund, 2003). The findings presented here suggest that giving can potentially be driven by guilt induced by prior unethical behavior, which speaks to recent research showing that non-standard motivations may impel prosocial behavior (Dana et al., 2007; Sachdeva et al., 2009).

Besides charity, the results also highlight the potential importance of guilt-based pricing for businesses. For example, travelers flying out of some airports receive the opportunity to offset the carbon footprint of their flight. Using "Climate Passport kiosks," people can calculate how many pounds of carbon dioxide their trip will produce and the cost of offsetting this footprint using donations to programs aimed at greenhouse gas reduction. Several online travel retailers have begun to offer a similar option—giving customers the choice of offsetting their carbon footprint directly after ticket purchase. This kind of business is in line with our hypotheses: people clear their bad feelings by donating. According to the conscience accounting hypothesis, programs that ask for donations close to the time of unethical purchase should be more successful than alternatives that ask people to donate at a more remote time or before the unethical purchase is made.

In addition, the dynamic effects of guilt on prosocial behavior adds to the growing literature on the "demand" side of philanthropy (DellaVigna et al., 2012).

Although the emotional response discussed here is temporary, it may be used strategically to increase prosocial acts by individuals and organizations wishing to maximize donations. Furthermore, reminders of past unethical actions might lead to similar emotional dynamics as outlined in this paper. People may want to avoid guilty feelings, but will still act more prosocially if reminded about the ethical dimensions of past or current actions. Similarly, reminders of past immoral choices—such as directing a person's attention to her broken promises or deceptions—may induce feelings of guilt that can facilitate greater donation revenue for charitable institutions and to increased loyalty within organizations.

Our proposed framework is also linked to the economic literature of incorporating procedural norms into economic behavior (Kahneman et al., 1986; Akerlof and Kranton, 2000; Cappelen et al., 2007; Kircher et al., 2011). Particularly, in our framework, individuals would prefer to adhere to procedural norms when attaining a given consumption vector. Upon violating a norm, however, they exhibit a temporal altruistic preference reversal toward others. In this manner it is possible to identify norm violations in observable behavior.

The rest of the paper is organized as follows. In Section 2 we introduce and develop our behavioral hypotheses. In Section 3 we present evidence from a deception game experiment in which we provide support for our predictions. Section 4 lays out the results of an "over-paying" experiment that provides further support for our theory, and discusses the related moral licensing hypothesis (Monin and Miller, 2001) in more depth. In Section 5, we discuss several examples of how conscience accounting can be utilized by firms to maximize revenue, and posit how our framework can be generalized to other emotions.

4.2 Behavioral Hypotheses

In this Section, we outline three behavioral predictions based on a simple formal model of emotional decision making in the presence of moral constraints.[2] Consider an individual who faces a temporal sequence of allocation decisions between

[2]To better focus on the key mechanisms, we present a very simple setting. For a more broadly applicable model and demonstration that the effects described here extend to general allocation problems, see Gneezy et al. (2012).

herself and others leading to a final consumption vector. In specifying the decision maker's preferences, we extend the basic model of altruism, e.g., Becker (1976). In particular, we assume that the decision maker internalizes a set of moral constraints or norms and experiences feelings of *guilt* upon violating these constraints. Similar to Akerlof and Kranton (2000), norms here describe what the individual should not do; rather than prohibiting specific payoffs, these constraints prohibit procedures and actions by which these payoffs are attained. Examples of such procedural norms include attaining the same payoff allocation by either lying or telling the truth, by stealing from business partners or receiving a gift, with the former procedure constituting a moral transgression in each case. Conditional on attaining the same payoff, individuals in our framework have a clear preference for not violating a moral constraint because doing so induces negative feelings of guilt.

To formalize this idea, we assume that the individual makes a sequence of allocation decisions in periods $t = 1, 2, 3$ the sum of which produces a final allocation vector to be consumed in period $T = 4$.[3] To incorporate emotions further, we take that in each decision period t, the individual derives anticipatory utility based on her expectation of the final consumption vector and her current feelings of guilt (Loewenstein, 1987). Specifically, the decision maker's anticipatory utility at time t depends on her emotional state experienced in that period, $g_t \in \mathbb{R}^+$, which is interpreted as the intensity of her guilt. Since consumption takes place in period $T = 4$, her anticipatory utility in a periods $t < T$ can be expressed as

$$u(x, y, g_t)$$

where x is her own final consumption in period T, y is the total consumption given to others in period T. In line with standard models of altruism, both x and y are taken to be normal goods and u to be twice differentiable.

In a context where the individual experiences multiple periods of anticipatory emotions, her preferences at time t are defined by maximizing the sum of current and future anticipatory utilities and final consumption utility under rational expectations. Thus the individual's preferences in period t are given by:

[3] The model directly extends to any finite number of periods T.

$$U_t = E_t \sum\nolimits_{s=t}^{s=4} u(x, y, g_s)$$

Note that if the level of guilt remains the same over time, then preferences over the final allocations are stable and do not change. In contrast, when the level of guilt changes from one period to another—which will always be the case after a moral transgression—her preferences over the final allocations may also change over time. Thus emotional fluctuations due to moral transgressions lead to time-inconsistency in behavior. We will exploit such time-inconsistency in identifying the implications of the model.[4]

4.2.1 Psychological Framework

We now turn to the framework characterizing how guilt affects preferences, first describing a simple dynamic of how guilt evolves over time, then discussing the relationship between guilt and preferences.

As is typical of many emotions, a class of events triggers a rapid and relatively large subsequent change in an individual's emotional state. With the passage of time, the individual "cools off" and the emotional state reverts back to baseline (Loewenstein, 1996; Elster, 1998; Van Boven et al., 2009). Describing the dynamics of guilt, we assume that the violation of a moral constraint is followed by a *subsequent* increase in feelings of guilt, and that this guilt gradually subsides over time. Formally,

$$g_{t+1} = \gamma g_t + z \text{ where}$$
$$z = 1 \text{ if choice at } t \text{ violates a moral constraint, } z = 0 \text{ otherwise.}$$

where $\gamma \in (0, 1)$ expresses the speed at which guilt decays.

In our setting, guilt affects utility in two ways: it is not only an aversive feeling, but one that also facilitates prosocial interactions (Baumeister et al., 1994; de Hooge, 2012). Guilt is costly and one would want to avoid experiencing it conditional on being able to attain the same payoff allocation. Additionally, guilt is accompanied

[4] As standard, given time-inconsistency and rational expectations, the solution of the decision maker's maximization problem is obtained via backward induction.

by the subsequent desire to treat others well and behave more prosocially (Keltner and Lerner, 2010). In turn, guilt can be characterized as a relative substitute of a person's own consumption: feelings of guilt increase the importance of improving the consumption of others relative to one's own. We assume that people prefer to avoid violating norms because of guilt, and conditional on violating a norm, behave more prosocially. Formally,

$$u_g < 0 \text{ and } u_{g,x} < 0 \leq u_{g,y}$$

4.2.2 Hypotheses

Below we derive three predictions of the model outlined above. Since emotions are inherently temporal, the identification of our proposed mechanism relies critically on the link between the timing of choice and the resulting behavior.

To test the main implications of the above model, it will suffice to consider two simple choice sets arranged over three periods. Let $M = \{(x_1, y_1), (x_2, y_2)\}$ be a choice set with two allocation options, where x refers to the decision-maker's own material payoff, y to the material payoff of others and assume that $x_1 > x_2$. Let $D = \{(-d, d), (0, 0)\}$ be a donation choice set between a positive transfer from the decision-maker to others (a donation) and an allocation $(0, 0)$ with no transfer (no donation). Finally, let the empty choice set $\emptyset$ simply describe a period where there is no allocation decision to be made. To render the problem non-trivial, unless mentioned otherwise, we assume that choosing the more selfish allocation from M requires a norm violation / moral transgression

We can now consider two distinct choice problems. Let $S_{\text{hot}} = \{M, D, \emptyset\}$ describe the choice problem where the donation option directly follows the potential norm violation, and let $S_{\text{cold}} = \{M, \emptyset, D\}$ describe the alternative choice problem where the donation option is presented after some time has passed. Given these choice problems we first describe a hypothesis that directly encapsulates conscience accounting.

In our framework, guilt dynamics imply that after a moral transgression people exhibit a greater willingness to act prosocially. Hence, in a temporal bracket following a potential norm violation we expect to see more prosocial behavior. As guilt decays with time, this increased tendency for prosocial behavior diminishes as well.

Claim 1 (**Conscience Accounting Hypothesis**). *The decision maker is more likely to donate after a norm violation if the donation option is presented earlier than if it is presented later, i.e. in* S_{hot} *versus in* S_{cold}. *If the donation option is known ex-ante, the decision maker is also less likely to jointly violate a norm and not donate in* S_{hot} *than in* S_{cold}.

Note that the conscience accounting hypothesis holds both if the donation option is a surprise as well as when it is known ex-ante. Furthermore, if the donation is a surprise, the propensity to donate after an immoral choice is higher in S_{hot} than in S_{cold}, while it is the same after a moral choice. We test these predictions directly in the next Section.[5]

To account for the fact that violating a moral constraint may lead to a higher earning for the decision maker (e.g. $x_1 > x_2$), we will also test whether the same increased tendency for prosocial behavior is observed when the decision maker's payoffs are kept constant but there is no possibility of a norm violation. Our mechanism suggests that if the decision maker attains the same payoff *without* violating a moral constraint then we should not observe a temporal increase in prosocial behavior. This prediction is discussed further when addressing potential income effects in Sections 3 and 4.

In the absence of knowing what acts violate norms and what acts do not, the above result also allows us to identify moral constraints in dynamic choice behavior. Holding the material payoffs constant, our first hypothesis links a specific temporal choice pattern to the presence of a moral violation. If an act violates a constraint, the level of subsequent altruism follows a predictable pattern—greatest after the act and returning back to baseline after some time has passed.

While the first hypothesis centered on the retrospective effects of guilt on behavior, the next two hypotheses focus on the prospective effects of guilt, i.e., how people respond to emotional fluctuations ex ante. The dynamic pattern of emotions

[5] In the case where the donation option is known ex-ante, the timing of the donation option may effect the initial decision to violate a norm. We discuss this below.

implies that as guilty feelings change, so do preferences over allocations. If a person knew that she would have an opportunity to donate after an unethical act, then being at least partially aware of the conscience accounting response can lead to a dynamic impulse control problem: the moral violation produces an altruistic urge that the individual would potentially want to control prior to the violation. If the individual believes that the subsequent guilt will induce her to be more charitable than she finds optimal ex ante, she would prefer to limit her ability to donate, e.g. by delaying the opportunity until she feels less guilty, limiting ex post donation options, or not violating the norm at all.[6]

At the same time, our assumption that guilt is a relative substitute of own consumption implies that there is always a general complementary relationship between moral violations and prosocial actions—moral violations increase one's desire to be prosocial. In particular, if a desirable allocation can only be attained by violating a norm, then this allocation is more likely to be chosen if a limited donation option is available shortly after the unethical act. "Paying for her sins" by donating after unethical behavior will make a guilt-prone individual feel better: since guilt is a relative substitute for own-consumption over the consumption of others, moving towards a more altruistic allocation will help reduce the overall utility loss from immoral behavior.

To demonstrate the latter effect, take the case where an individual always prefers the selfish allocation when it does not require a moral transgression, but prefers the less selfish one when it does —finding the utility cost of guilt associated with deception to outweigh the payoff difference. Since guilt is a bad in our model, the required moral transgression leads the individual to choose the selfish allocation less often than when immoral actions are not required to attain it.[7] Here, adding the option to donate after the potential norm violation will allow the decision-maker to

[6] DellaVigna et al. (2012) provide evidence that the ex-ante ability to avoid later donation solicitations reduces donations by 28%-44%. Andreoni et al. (2011) provide further evidence for such avoidance techniques.

[7] Such preferences are consistent with Gneezy (2005), who demonstrates that individuals are significantly more averse to obtaining a selfish allocation through unethical behavior such as lying than choosing it from a set of allocations, e.g. in a Dictator game. For example, 90% of individuals chose $(15, 5)$ over $(5, 15)$ in a Dictator game while only 52% were willing to obtain the $(15, 5)$ allocation when it required a lie ($p < .01$).

'lower' the overall utility cost of transgressing, making her more likely to transgress and subsequently donate. In turn, informing the decision-maker of a limited prosocial opportunity will increase the likelihood of norm violation, and if she does violate a norm, the decision-maker will also choose to subsequently donate.

To state the second hypothesis, consider a choice problem where there is no donation option: $S = \{M, \emptyset, \emptyset\}$.

Claim 2 (**Paying for Sins Hypothesis**)**.** *Suppose that the decision maker does not violate the norm in S. Informing the decision maker about a future donation option ex-ante, i.e., that she is facing S_{hot}, will increase the likelihood of a norm violation, and if she now transgresses, she will also donate in S_{hot}.*

Given an ex-ante choice to be presented with a binary donation option either soon after a moral transgression or after some time has passed, one can demonstrate the interplay between the demand for altruism and the worry of paying too much for one's sins. Since guilt diminishes over time, a decision to delay the donation option can serve as a commitment to give less. Similarly, a decision to be presented with the donation option early can serve as a commitment to give more. People who morally transgress may exhibit both preferences. It is always true, however, that those who find it optimal to "pay for their sins" would strictly prefer the early opportunity—because they fear that their later and colder self will not be generous enough—and donate. On the other hand, those who prefer not to donate and are worried about the above discussed impulse control problem will strictly prefer the late donation opportunity, and then will not donate. This link between donation behavior and preference for timing leads to our third hypothesis:

Claim 3 (**Choice Hypothesis**)**.** *When presented with the ex-ante choice of either facing an early or a late donation option, S_{hot} or S_{cold}, those who strictly prefer an early opportunity and violate the norm will donate. Those who strictly prefer a late opportunity and violate the norm will not donate.*

Above, we described how varying the temporal distance between choices or the information about this distance could be used to identify guilt dynamics. In addition, the proposed mechanism speaks to how behavior might change when varying the *order* of decisions. Particularly, consider the comparison between choice sets where the donation option precedes the potential moral violation, $S_{\text{pre}} = \{D, M, \emptyset\}$, and the case where it follows, S_{hot}.

In this context, one can state an *alternate hypothesis* to conscience accounting where people simply alternate between being prosocial or not, but are not subject to guilt dynamics. Under the alternate hypothesis, the ordering of choices should not matter with regards to the overall pattern of behavior. Specifically, the proportion of individuals who choose to donate and then lie and those who do not donate and tell the truth should be similar whether the donation option follows the decision to lie or precedes it. If lying positively predicts donations in S_{hot}, then under this alternative hypothesis, donating should predict lying to a similar extent in S_{pre}.

In contrast, the presence of guilt dynamics under the conscience accounting hypothesis suggests a different prediction. Since guilt increases prosocial motivation, there should be more donations by those who violate a norm when the donation option follows the potentially immoral choice than when the ordering is reversed. This prediction follows from our model because guilt increases only *after* a moral transgression. Hence, there should be less individuals who jointly lie and do not donate in S_{hot}, which suggests a significant correlation between the immoral choice and donations in S_{hot} but not in S_{pre}. The predicted difference in correlations rests on the assumption that not donating is considered to be less of an unethical act than the moral transgression in M, e.g., lying or stealing. We present empirical support for this assumption in the next section.

4.3 A Deception Game

4.3.1 Procedure

To study conscience accounting empirically, we conducted a two-stage experiment. First, participants could lie to increase their profits at the expense of another participant. Second, after choosing whether to lie, we gave participants the option to donate to a charity.

We used a setup similar to Gneezy (2005). In this two-player deception game, one player, the Sender, has private information and the other, the Receiver, makes a choice based on a message conveyed by the Sender. The payoffs for both players depend on the choice the Receiver makes. We constructed payoffs such that lying (sending a "wrong" message) resulted in a higher payoff for the Sender.

In the instructions (see Online Appendix)[8], we told participants that the experiment had two possible payment outcomes. Although the Receiver's choice would determine the outcome, only the Sender knew about the monetary outcomes of each option—the Receiver had no information regarding the alignment of incentives. As was shown using this game (Gneezy, 2005; Sutter, 2009; Dreber et al., 2008), most Receivers choose to follow the message sent by the Sender, and most Senders expect this.

After choosing whether to lie, Senders were given the option to donate to a charitable foundation. Depending on the treatment, the donation option was either presented directly after the decision to lie or with some delay, and Senders were either aware of this option when choosing whether to lie or not.

We recruited 528 undergraduate students at the University of California, San Diego for a classroom experiment. The rules of the experiment were both read aloud and presented in written form to the participants.

We informed them that neither the Sender nor Receiver would ever know the identity of the player with whom they were matched. Participants in both roles knew that 1 of 10 students assigned the role of Sender would be randomly chosen to be paid, and be matched with Receivers in a different class.

Senders could choose from one of 10 possible messages to send the Receiver. Each message was in the form of "Choosing _ will earn you more money than any other number," with the blank corresponding to a number from 0 to 9. We told the Sender that if the Receiver chose a number that corresponded to the last digit of the Sender's Personal Identification number (PID), both players would be paid according to payment Option Y, and if the Receiver chose any other number, both players would be paid according to Option X. We informed Senders of the monetary consequences of both Option X and Y, and that the Receivers were not informed of this. We constructed the payments such that Option Y earned the Receiver more

[8] https://sites.google.com/site/alexoimas/ca

money than the Sender, and Option X earned the Sender more money than the Receiver. Hence, if the Sender expected the Receiver to follow her message, she had a monetary incentive to send one that did not correspond to the last digit of her PID—to lie—so the Receiver would choose the wrong number. Gneezy (2005) showed that Senders in the deception game expected the Receivers to follow their message (82%), and that Receivers indeed largely followed the message sent (78%). Receivers in our experiment largely followed the Senders' messages as well, with 75% of participants assigned the role of Receiver choosing the number the Sender wrote would earn them the most money.

All treatments (other than *Baseline*) offered Senders the option to donate either \$1 or \$2 to the Make-A-Wish foundation. These numbers were used so that the donation amount would always be relatively small in comparison to the amount that could be gained through deception. In the *Incentive* and *No Incentive* treatments, we presented the donation option directly after Senders made their message choices. In the *Incentive Delay* treatment, we presented the donation option with some delay: after their message choice, Senders received anagrams to solve for 10 minutes before we presented them with the option to donate. Importantly, in these three treatments Senders were not aware of the subsequent donation option when choosing what message to send, but were informed of it only after they made their initial choice. The *Incentive Reverse* treatment was similar to the Incentive treatment, except that the donation option was presented before the message choice. Senders were not aware of the deception game when making the donation decision.

In the *Informed Incentive* and *Incentive Choice* treatments, however, Senders knew in advance they would have the opportunity to donate. Other than Senders being informed that this opportunity would directly follow their message choice (and the different payoffs), the Informed Incentive treatment was similar to the Incentive treatment. In the Incentive Choice treatment, we asked Senders to choose whether they wanted to make the decision to donate sooner (directly after their message choice) or later (at the end of the experiment), at the same time as deciding what message to send. Senders made the actual donation decision according to this choice. Ten minutes of anagrams once again served as the delay.

The last treatment was a baseline containing the same payoffs as the Informed

Incentive treatment but excluding the donation option.

Table 4.1 presents the payoffs we used in the experiment. We designed the Incentive, Incentive Delay, Incentive Reverse and Incentive Choice treatments such that if the Receiver chose the wrong number, the Sender stood to earn $10 more and the Receiver $10 less than if the Receiver chose the correct number. Senders had a smaller incentive to lie in the Informed Incentive and Baseline treatments, earning $5 more and the Receiver $5 less if the Receiver chose the wrong number.

In the No Incentive treatment the Sender had no monetary incentive to lie: both the Sender and Receiver stood to potentially earn $10 less if the Receiver chose the wrong number. Note that a Sender in the No Incentive treatment could obtain the same payoff by telling the truth as she could obtain by lying in the Incentive treatment.

Since Receivers were not informed about the Senders' payoffs, we did not expect nor did we observe a difference in the Receivers' behavior.

We established subject identification through the PID numbers students provided as part of the experiment. We used the PID numbers to pay the participants according to the outcome of the experiment and to determine whether the Sender had lied in her message. Donations were either $1 or $2 in each available case, and we deducted the amount from the Senders' payments if they chose to donate. We then made the donations on behalf of the Senders directly through the Make-A-Wish website. Using these treatments we test our hypotheses.

Conscience Accounting Hypothesis: The Incentive and Incentive Delay treatments allow us to test the Conscience Accounting hypothesis. Conscience accounting predicts that those who lied in the Incentive treatment should donate to a greater extent than those who lied in the Incentive Delay treatment. Furthermore, we should also observe a greater correlation between the decision to lie and the decision to donate in the Incentive treatment than in the Incentive Delay treatment.

In addition, comparing the Incentive and Incentive Reverse treatments, conscience accounting implies that there should be a greater correlation between lying and donating when the decisions follow in that order (Incentive treatment), than when the donation option precedes the decision to lie (Incentive Reverse treatment). Particularly, the decision to lie or not should predict the decision to donate in the former treatment but not the latter.

We use the No Incentive treatment to test whether the moral character of a choice has an impact on prosocial behavior after controlling for earnings. The prediction on donation rates in the Incentive and No Incentive treatments follows from our model, and allows us to rule out that conscience accounting is driven by differences in the Senders' income. Those who lie in the Incentive treatment stand to earn the same amount as those who tell the truth in the No Incentive treatment ($20). Our framework predicts that those who tell the truth to earn $20 in the No Incentive treatment should feel less guilty, and in turn, be less likely to donate than those who lie to earn $20 in the Incentive treatment.

Paying for Sins Hypothesis: The Informed Incentive and Baseline treatments were designed to test the Paying for Sins hypothesis. Given the small size of the donation, the hypothesis would predict that those in the Informed Incentive treatment should be more likely to lie than those in the Baseline treatment, where Senders were not informed of a subsequent donation opportunity. This prediction holds under the assumption that more individuals would prefer the selfish allocation in the Baseline treatment if they did not have to lie to attain it (for evidence that this assumption holds, see Gneezy, 2005).

Choice Hypothesis: The Incentive Choice treatment allowed us to test the Choice hypothesis which predicts that those who lied and chose the early donation opportunity should donate more than those who lied and chose the late opportunity.

Is Lying a Moral Violation? To provide external evidence in support of the moral category assumptions in the experiment, we adopted the incentive-compatible elicitation method of Krupka and Weber (2013) to assess people's moral attitudes towards the various choice options in our paradigm. We presented a separate group of participants with a description of the deception game and the possible choices available to the Sender, and used the method to elicit whether they viewed each choice as morally appropriate or inappropriate. Participants were instructed to view a morally inappropriate choice as one that, if chosen, would make the acting individual feel guilty. To incentivize truthful reporting, participants were paid an additional fee if their response to the moral appropriateness of a choice matched the response selected by most other people in the experiment. In turn, each was prompted to coordinate on the belief held by the social group as to whether a choice violates a moral constraint or not.

Participants (N=43) were recruited using the Amazon Mechanical Turk platform. Each read the Sender's instructions for the deception game and was asked to judge the extent to which each of the Sender's possible message choices was "morally appropriate" and "consistent with moral or proper social behavior" or "morally inappropriate" and "inconsistent with moral or proper behavior." Participants were told that at the end of the experiment, we would randomly select one of their choices, and if their moral appropriateness rating matched the modal response of others in the experiment, they would be paid an additional \$3 on top of a base fee of \$0.50. If their rating did not match the modal response, no extra payment would be awarded.[9]

To test the assumption that donating nothing is not viewed as a moral violation, we separately elicited the moral appropriateness ratings for the choice to donate \$2 from experimental earnings to charity as well as the choice *not* to donate.

4.3.2 Results

Moral Attitudes

Following Krupka and Weber (2013), we converted participants' responses into numerical scores.[10] See Online Appendix for summary ratings of all evaluated choices as well as robustness checks of the analysis.

Looking at message choices that did not match the Sender's personal code, the average rating given was -.40. Pairwise t-test comparisons revealed no significant differences between messages that we classify as lies—each was judged to be a moral violation if chosen by the Sender (all p-values $> .5$). On the other hand, the message matching the Sender's personal code had an average rating of .71. In pairwise comparisons the message that we classify as telling the truth was rated significantly more morally appropriate than any of the messages classified as lies (all p-values $< .001$).

[9]See Online Appendix for instructions

[10]Participants chose from amongst five categories and numerical scores were assigned such that "very morally inappropriate"=-1, "somewhat morally inappropriate"=-1/3, "somewhat morally appropriate"=1/3, "morally appropriate"=2/3 and "very morally appropriate"=1.

Looking at donation decisions, the decision to donate had an average rating of .66: significantly more morally appropriate than any of the decisions to lie (all p-values $< .001$). Importantly, the decision *not* to donate, with an average rating of .09, was also judged to be significantly less of a moral violation than any of the choices to lie (all p-values $< .01$). In turn, we find support for our assumption that the decision not to donate does not constitute a moral violation while the decision to lie does. In light of our theory, this classification is consistent with the behavior revealed in the experiments described below.

Behavioral Hypotheses

Lying rates by treatment are presented in Table 4.1. The differences in lying rates between the Incentive and Incentive Delay treatments ($Z = 1.02,\ p = .15$), Incentive and Incentive Reverse ($Z = .47,\ p = .32$) the Incentive and Incentive Choice treatments ($Z = .43,\ p = .33$) were not statistically significant.[11] However, this difference between the Incentive and No Incentive treatments ($Z = 4.32,\ p < .001$) was statistically significant.

To test our first hypothesis, we rely on identification through time by examining behavior in the Incentive and Incentive Delay treatments. Our first key finding is that in the Incentive treatment, when the donation option came as a surprise directly after the message choice, 30% (6) of the participants who told the truth chose to donate, compared to 73% (27) of those who lied ($Z = 3.14,\ p < .001$): the participants who chose to lie—and potentially earn $10 from lying—were significantly more likely to donate to charity than those who chose to tell the truth. Note that this finding demonstrates that classifying individuals into simple "types," where some always behave in a "moral" way (not lying and donating) and others never do (lying and not donating), is problematic. In our experiment, those who had previously lied were *also* more likely to donate to charity.

In the Incentive Delay treatment, where the option to donate was presented some time after the message choice, 33% (3) of the participants who sent a true message chose to donate compared to 52% (14) of those who lied ($Z = .96,\ p = .17$).

[11] p-values were calculated from a one-tailed test of the equality of proportions using a normal approximation to the binomial distribution.

Looking at the conditional propensities to donate in the Incentive and Incentive Delay treatments, those who lied and had the opportunity to donate directly after their message choice did so significantly more often than those who lied and faced a delay between the two choices ($Z = 1.74,\ p = .04$). On the other hand, the delay had *no effect* on the donation rates of truth tellers ($Z = .18,\ p = .43$). These results are summarized in Figure 4.1.

Since the only difference between the Incentive and Incentive Delay treatments was the time between the message choice and donation option, the observed difference in donation rates of liars also rules out a number of potential alternate explanations. Particularly, if the higher donation rate of liars compared to truth tellers in the Incentive treatment was due to Senders having heterogeneous preferences over allocations between Receivers, themselves, and the charity, then Senders who lied should have donated at the same rate in both treatments. An explanation based on some form of self-image preservation—of trying to cancel a bad deed with a good one—would also predict that donation rates should not change between treatments. In addition, since those who lied in both Incentive and Incentive Delay treatments were expecting to earn the same higher payoff of $20, the greater donation rates of liars in the Incentive treatment cannot be explained by differences in earnings.

Looking at the relationship between lying and donating, an OLS regression revealed that the decision to lie had a significant influence on the decision to donate in the Incentive treatment ($\beta = .401,\ p = .001$). However, in line with our prediction, there was no relationship between the decision to lie and donating in the Incentive Reverse treatment where the order of the decisions was reversed ($\beta = .02,\ p = .87$). Regressing the choice to donate on the decision to lie, a treatment dummy, and the interaction of the decision to lie and the treatment dummy revealed a significant interaction effect ($\beta = -.404,\ p = .041$). The positive relationship between lying and donating was significantly greater when the donation option followed the decision to lie than if the choices were reversed. Individuals chose to lie and donate less in the Incentive Reverse treatment than in the Incentive treatment.

The difference in the relationship between lying and donating in the Incentive and Incentive Reverse treatments provides additional support for our conscience accounting hypothesis. If the increased donation rate of liars in the Incentive treatment was due to individuals broadly bracketing moral choices, offsetting a lie with a donation or vise versa, then the ordering of choices should not effect observed behavior. However, guilt dynamics imply that the positive relationship between lying and donating should be larger when the donation option *follows* the decision to lie and not when the order of choices is reversed. Our results are consistent with the latter prediction.

To provide further support that differences in donation rates between liars and truth tellers in the Incentive treatment are not due to differences in the Senders' material payoffs – an income effect – we compare the donation rates of Senders whose material payoffs are held constant but who differ in whether these payoffs were attained by lying or telling the truth. If the higher donation rates of liars were caused by greater material payoffs in the deception game, then the same higher donation rates should be observed when Senders were truthful in the No Incentive treatment. In the No Incentive treatment, Senders did not have a monetary incentive to lie. Specifically, their expected payoffs from truth telling in the No Incentive treatment was the same as from lying in the Incentive treatment: \$20. Hence, the only difference between earning \$20 in the Incentive and No Incentive treatments is that a moral violation is required in the former but not the latter.

Consistent with our hypothesis, we found that in the No Incentive treatment, of those who told the truth, 51% (21) chose to donate compared to 73% (27) of those who lied in the Incentive treatment. Those who lied in the Incentive treatment were still significantly more likely to donate than those who had told the truth in the No Incentive treatment ($Z = 1.97,\ p = .02$), despite the fact that the expected own payoffs were the same.

Combined, these results provide direct support for the Conscience Accounting hypothesis which predicts that when the subsequent donation option comes as a surprise, individuals who violated a norm will be more likely to donate than those who did not within a temporal bracket.

We now turn to our predictions for contexts where Senders are informed of the donation option in advance. To test our Paying for Sins hypothesis, we compare the lying rates of Senders in the Informed Incentive treatment to those in Baseline. In line with our predictions, 63% (39) of Senders lied in the Informed Incentive treatment compared to 48% (30) of those in the Baseline treatment($Z = 1.72,\ p = .04$).[12] Of those who lied, 82% donated. Given the relatively low rate of deception in the Baseline treatment, these results provide support for the Paying for Sins hypothesis: knowing that they would be presented with the opportunity to "pay for their sins" in the future led to Senders more willing to lie.[13]

In the Incentive Choice treatment, where Senders could choose when to be presented with the donation option, of those who lied 43% (10) chose to make their donation decisions early and 57% (13) chose to make their donation decisions late. Of those who chose to make their donation decision early and lied, 90% (9) actually donated, compared to 31% (4) of those who chose to decide later and lied ($Z = 2.84,\ p < .001$). These results provide support for the Choice Hypothesis: when given the choice of when to be presented with the donation option, those who chose to be presented with it early donated much more than those who chose to be presented with it late.[14]

We did not find any significant gender differences in behavior.

4.4 An Over-paying Experiment

4.4.1 Procedure

In the deception game experiment, participants knew that we were observing whether they had lied. We designed the second experiment such that participants were unaware we were studying their moral choices. This unawareness should reduce

[12] To test the robustness of these results, we ran these treatments again with a different group of subjects using the same instructions ($Z = 1.74,\ p = .04$). Combining results across both iterations yielded similarly significant results ($Z = 2.02,\ p = .02$).

[13] Although a direct comparison to prior research is not appropriate, we note that Gneezy (2005) observed 66% of individuals chose the selfish allocation in a Dictator game for a gain of $1, while we observed that only 48% of Senders were willing to lie to obtain a larger gain of $5 in the Baseline treatment. This suggests that Senders would be more willing to choose ($20, $15) over ($15, $20) if it did not require a lie.

[14] It should be noted that there was no increase in deception in the Incentive Choice treatment when compared to the Incentive and Incentive Delay treatments where individuals were not informed of the subsequent donation option ($Z = .91,\ p = .18$). We believe that this was due to the fact that in addition to being informed of the subsequent donation option, individuals in the Incentive Choice treatment were also asked to make a choice on the timing of this option, which may have interacted with their decision to lie.

behavior based on the experimenter demand effect and/or experimenter scrutiny.

We paid groups of subjects for their participation in an unrelated experiment. Two groups received payment according to how much we promised them. A third group received more than they were promised by "mistake" and had the opportunity to either return or keep the extra money.[15] We also gave all three groups the option to donate $1, and we recorded donation rates across the groups. According to our hypothesis, we expected conscience accounting to manifest itself in the third group, predicting participants who decided to keep the extra money for themselves would be more likely to donate than those who had returned it.

We recruited 160 undergraduate students at the University of California, San Diego to participate in a coordination game experiment with an advertised expected payoff of $10 (Blume and Gneezy, 2010). We invited subjects to the lab in pairs and seated them far apart for the duration of the game, which took approximately 15 minutes. We guaranteed all participants a $5 show-up fee, and those who did not succeed in coordinating did not get any extra money.

In addition, participants received $10 or $14, depending on the treatment, if they were able to coordinate with the individuals whom they were matched with. We randomly assigned those who had succeeded in coordinating to one of three treatments. In the Low treatment, subjects learned they would receive an additional $10 if they had succeeded in coordinating with their partners. In the High treatment, participants learned the additional payment would be $14. In the Mistake treatment, we told participants they would get $10 if they had succeeded, but gave them $10 and an extra $4 by "mistake"–nine $1 bills and one $5 bill interspersed amongst them. Table 4.2 summarizes payments for all three treatments. After receiving their pay at the end of the experiment, participants in all three treatments were given a description of a child with cancer and asked if they wanted to donate $1 from their final payment to the child.

When they received their pay participants were told, "Here is your _. Please count it and sign this form," with the blank corresponding to the promised payment

[15]The study of individuals who do not know they are participating in an experiment is a common practice in field experiments, see gift exchange experiments e.g. Gneezy and List (2006), and is used in part to minimize experimenter demand effects that may be present in the lab.

($10 in the Low and Mistake treatments, $14 dollars in the High treatment). Then the experimenter left the room. All payments were made in $1 bills except for the extra $5 bill in the Mistake treatment. Participants in all three treatments then decided whether to donate.

This framework allowed us to test the Conscience Accounting hypothesis in a different context. Particularly, we expected participants in the Mistake treatment who did not return the extra money to be more likely to donate than those who had returned it. In addition, since attaining $14 in the Mistake treatment required a moral transgression while earning $14 in the High treatment did not, we predicted that donation rates in the former treatment would be significantly greater than in the latter.

To support that our moral category assumptions were met, we again used the method of Krupka and Weber (2013) to elicit moral judgements about the choices participants faced in this paradigm. Particularly, we gave the same group of participants as in the first study an exact description of the over-paying experiment. We presented them with the choices available to an individual in the Mistake treatment, and asked each to judge the extent to which keeping the extra money and returning the extra money was morally appropriate.

4.4.2 Results

Moral Attitudes

As in the first study, we converted participants' responses into numerical scores. Looking at the choice of not returning the $4 given by mistake, the average rating given was -.67. On the other hand, returning the $4 given by mistake was given an average rating of .86 ($t(84) = 23.82$, $p < .001$). In line with our assumptions, keeping the extra money was viewed as a moral violation while returning the extra money was viewed as morally appropriate.

Behavioral Hypothesis

In the Mistake treatment, 41% (33) of participants returned the extra money they had received by "mistake." Donation rates by treatment are presented in Figure

4.2. Overall donation rates of participants were 30% (12) in the Low treatment, 25% (10) in the High treatment, and 49% (39) in the Mistake treatment. The overall donation rate in the Mistake treatment was significantly higher than in both the Low ($Z = 1.96,\ p = .03$) and the High ($Z = 2.50,\ p = .01$) treatments. Consistent with the Conscience Accounting hypothesis, of those who returned the extra money in the Mistake treatment, 27% (9) made a donation, whereas 64% (30) of those who did not return the extra money made a donation ($Z = 3.22,\ p < .001$).

An income effect of earning $14 rather than $10 does not explain the discrepancy in donation rates. Subjects in the High treatment, who were told they would earn $14,[16] donated at about the same rate as those who returned the extra money but significantly less than those who kept it. Namely, although the donation rate for participants who returned the extra money is similar to those in the Low ($Z = .17,\ p = .43$) and High ($Z = .22,\ p = .41$) treatments, the donation rate for those who kept the money is significantly higher ($Z = 3.15,\ p < .001$ and $Z = 3.62,\ p < .001$, respectively). The difference in behavior in the Mistake treatment also suggests many participants, including those who did not return the money, did notice the mistake.

Our results also speak to the "moral licensing" hypothesis proposed by Monin and Miller (2001), where past moral actions can justify less moral choices down the road (Khan and Dhar, 2006; Zhong et al., 2009). For example, Monin and Miller (2001) showed that participants allowed to establish themselves as not being prejudiced were more likely to later make remarks deemed socially offensive. One way to interpret moral licensing in the context of the over-paying experiment is to say that people who behaved morally and returned the extra money should be less likely to subsequently choose to donate than those achieving the same payoff without a moral act because they had earned the "license" not to. The results presented in Figure 4.2 reject this prediction: people who returned the extra money, and hence did not violate a norm, donated at the same rate as those who had no option to make such

[16]Note that all participants were recruited to participate in the study for $10. Hence in this case, as well as in the case of the mistake, the extra $4 could be treated as a windfall because participants were initially expecting only $10. Particularly, since the reference point coming into the study was the same in all treatments – participants had the same payoff expectations of earning $10 – we believe that a windfall argument is not sufficient to explain the differences in donation rates.

a moral choice.

In addition, looking at the deception game, we can compare the lying rates of Senders in the Incentive Reverse and Incentive treatments. Participants in the Incentive Reverse treatment were given an opportunity to donate before making their decision lie or tell the truth, while those in the Incentive treatment were not. Since not donating was not viewed as an immoral act, moral licensing would predict that giving individuals the opportunity to make a donation should increase the amount of subsequent deception (conditional on at least some individuals donating). We did not observe this effect: 65% of Senders lied in the Incentive treatment while 60% of Senders lied in the Incentive Reverse. Although the difference was not significant, directionally individuals appeared to lie less when they had a prior opportunity to donate.

It should be noted that an important feature of studies demonstrating licensing is that the initial prosocial act was costless to the subject. For example, the subjects in the Monin and Miller (2001) study had the opportunity to establish themselves as unprejudiced at no cost to themselves. In another example, Khan and Dhar (2006) demonstrated licensing by having a group of individuals engage in one of two hypothetical volunteer assignments; they were then more likely than controls to choose a luxury item over a necessary item. However, a recent study by Gneezy et al. (2011a) found that cost is a critical factor in licensing, showing that when the initial prosocial act came at a cost to the subject (as in our experiment), the licensing effect disappeared.

4.5 Discussion and Conclusion

We examine emotional dynamics in the context of social behavior. We posit and test several behavioral hypotheses where individuals care about the procedural aspects of their choices and, upon violating a norm, exhibit a specific time-inconsistency in their attitude towards others, preferring a more prosocial allocation after a norm violation than in its absence. This suggests an additional explanation for prosocial behavior in the presence of moral constraints: people donate to account for their conscience after making a morally bad choice. The fact that people who lie are more likely to donate to charity than people who tell the (costly) truth may seem counter intuitive. One goal of this paper is to reshape this intuition.

Throughout the paper we have focused on the specific emotion of guilt. However, other negative emotions such as anger may fit a very similar retrospective temporal pattern in the context of social behavior. While guilt changes preferences to be more prosocial, events that provoke anger affect preferences so that hurting the other party becomes more desirable within a temporal bracket (Card and Dahl, 2011). Angry individuals may lash out at others even at a cost to themselves if such an opportunity arises soon after a trigger, but may prefer to control this impulse ex ante. In this manner, anger functions as a temporal shock to choices directed against the payoff of others. Such effects of anger on decision making are greater immediately after the incitement than after some delay—consistent with the folk wisdom of anger management: "count to 10 before reacting."

Incorporating emotional dynamics into models of charitable giving and prosocial behavior would provide further insight for theory that aims to better understand both the incidence of altruism and norm violations. For example, in moral choices such as the decision to lie, research has shown that people have a cost associated with breaking internal moral constraints that manifests itself as a conditional aversion to lying (Dreber et al., 2008; Gneezy, 2005; Sutter, 2009). Our findings suggest that feelings of guilt may play a role in those costs.

Our paper also sheds light on how businesses may optimally bundle products when pricing goods whose purchase violates internalized moral constraints. Climate Passport kiosks—as mentioned earlier—or check-out donations in liquor stores might speak to the existence of these practices. By offering such bundles in close temporal proximity (in a bracketing sense), businesses may not only increase the propensity to spend on the prosocial activity, but simultaneously increase the likelihood that individuals choose the products which may violate their moral constraints.

Additionally, the general relationship between emotional brackets and decision making outlined in our predictions provides an important avenue for future research, both on how emotions affect identifiable economic choices and the ways in which these effects are used strategically by individuals and organizations.

Chapter 4, in part, has been submitted for publication of the material as it may appear in Management Science, 2014, Gneezy, Uri; Imas, Alex O.; Madarász, Kristóf, INFORMS, 2014. The dissertation author was the co-primary investigator and author of this paper.

4.6 Appendix

In this Section, we present the three hypothesis formally. To see that the results hold more generally for convex D and M sets—as well as for further results—please consult Gneezy et al. (2012).

Proposition 1 (Hypothesis 1). *Suppose the donation option is a surprise. Given a norm violation, the DM is more likely to donate in S_{hot} than in S_{cold}. Suppose the donation option is known ex-ante. The DM is less likely to jointly act immoral and also not donate in S_{hot} than in S_{cold}.*

Proof. Consider first the problem where the donation option is a surprise. Since the availability of a donation option is not known at $t = 1$, the perceived problems in $t = 1$ are identical. Hence the initial choices in S_{hot} and in S_{cold} are the same.

Suppose that there is no norm violation at $t = 1$, then there is no increase in guilt—$g_2 = g_1 = 0$—and hence continuation behaviors in the two choice problems S_{hot} and S_{cold} are also identical.

Suppose now that there is a norm violation at $t = 1$. This implies that $g_2 > 0$. To prove the above claim, we need to show that the DM will be more altruistic at $t = 3$ than at $t = 2$. To show this, we need to compare the marginal rate of substitution (MRS) between own consumption and the consumption of others at $t = 2$ versus at $t = 3$:

$$\begin{aligned} MRS_{t=2}(x,y) &= \frac{u_x(x,y,1)+u_x(x,y,\gamma)}{u_y(\pi,1)+u_y(\pi,\gamma)} \leq \frac{u_x(x,y,\gamma)+u_x(x,y,\gamma)}{u_y(x,y,1)+u_y(x,y,\gamma)} \\ &\leq \frac{u_x(x,y,\gamma)+u_x(x,y,\gamma)}{u_y(x,y,\gamma)+u_y(x,y,\gamma)} = MRS_{t=3}(x,y) \end{aligned}$$

where the inequality follows from the assumptions that $u_{x,g} < 0 \leq u_{y,g}$ and that $\gamma \leq 1$. Thus whenever the decision-maker prefers to donate at $t = 3$—in the cold problem S_{cold}—she will do so at $t = 2$—in the hot problem S_{hot}—as well.

Consider now the case when the DM is informed in-advance about the donation option. This can have two effects on the $t = 1$ choice, (i) encourage norm violations by the ability of donating later or (ii) deter norm violations by fear of donating later. Note however that effect (i) is always greater in S_{hot} than in S_{cold} given the above inequality, and effect (ii) is also always greater in S_{hot} than in S_{cold} for the same reason. Hence it follows that those who transgress in S_{hot} but not in S_{cold} will always donate in S_{hot}. Similarly, those who transgress only in S_{cold} but not in S_{hot}, will not donate in S_{hot}. Hence it follows that for any given pool of subjects the likelihood of both a norm violation and no donation is always greater in S_{cold} than in S_{hot}, i.e., $\Pr((x_1, y_1), \text{"no donation"} \mid S_{\text{hot}}) < \Pr((x_1, y_1), \text{"no donation"} \mid S_{\text{cold}})$. Q.E.D.

Proposition 2 (Hypothesis 2). *Suppose the DM chooses (x_2, y_2) in S. Suppose in S_{hot} the donation option is known ex-ante. She is then more likely to choose (x_1, y_1) in S_{hot} than in S, and if she does so, she will donate as well.*

Proof. Note first that since $u_{g,y} > u_{g,x}$, conditional on a norm violation, planning to donate at $t = 2$ can improve the decision maker's overall utility at $t = 1$. In other words, at $t = 1$, the DM can prefer to violate a norm and donate, to not violating the norm and not donating, while at the same time prefer the latter to simply violating the norm. Formally,

$$\sum_{t=1}^{4} u(x_1 - d, y_1 + d, g_t) \geq \sum_{t=1}^{4} u(x_2, y_2, 0) \geq \sum_{t=1}^{4} u(x_1, y_1, g_t)$$

Furthermore, since conditional on a norm violation, $g_2 > g_1$, if the DM at $t = 1$ would prefer to donate, given that $MRS_{t=2} < MRS_{t=1}$, she will have the same preference at $t = 2$. Hence if she she violates the norm in S_{hot}, but not in S, she must also donate in S_{hot}. Q.E.D.

Proposition 3 (Hypothesis 3). *Suppose that the donation option is known ex-ante. If the DM ex-ante strictly prefers S_{hot} to S_{cold}, then she will choose to donate in S_{hot}. If the DM ex-ante strictly prefers S_{cold} to S_{hot}, then she will choose not to donate in S_{cold}.*

Proof. Note first that since the set of feasible options are the same, an ex-ante strict preference for one choice problem over another can arise only if the choice behavior in the two problems differ. It then follows that a strict preference can only arise when there is dynamic inconsistency in at least one of the choice problems, i.e., a norm violation occurs in M.

Given an initial norm violation from M, it follows from the proof of Hypothesis 1 that whenever "no donation" is preferred at stage $t = 2$, it must also be preferred at stage $t = 3$. For the same reason, whenever a "donation" is preferred at stage $t = 3$, it is also preferred at stage $t = 2$. Hence the difference in final allocation choices can only arise when "donation" is only implementable in S_{hot} but not in S_{cold} or when "no donation" is implementable in S_{cold} but not in S_{hot}. As a consequence, a strict preference for S_{hot} over S_{cold} implies a strict preference for a "donation" from D. Similarly, a strict preference for S_{cold} over S_{hot} implies a strict preference for "no donation" from D. This implies that if the decision maker strictly prefers S_{hot} over S_{cold} , she will donate, and if she strictly prefers S_{cold} over S_{hot}, she will not donate. Q.E.D.

4.7 Figures

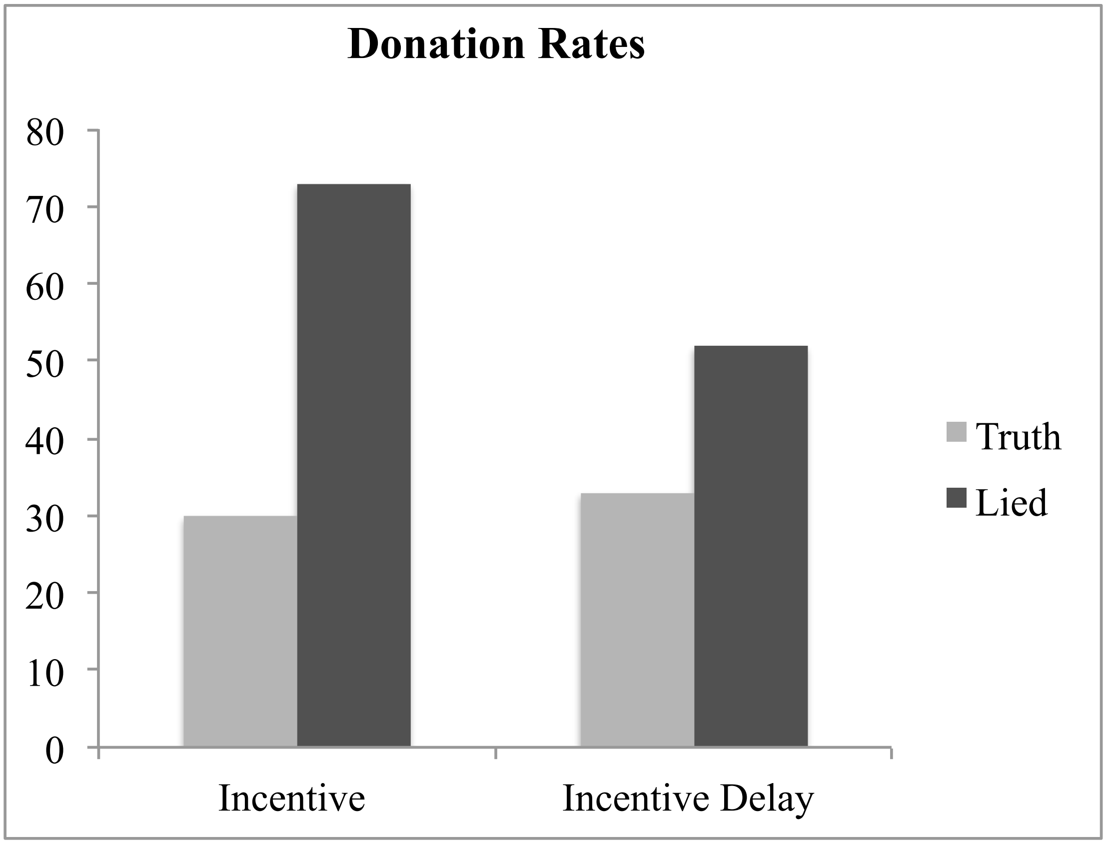


Figure 4.1: Fraction of Senders Who Donated by Message Type

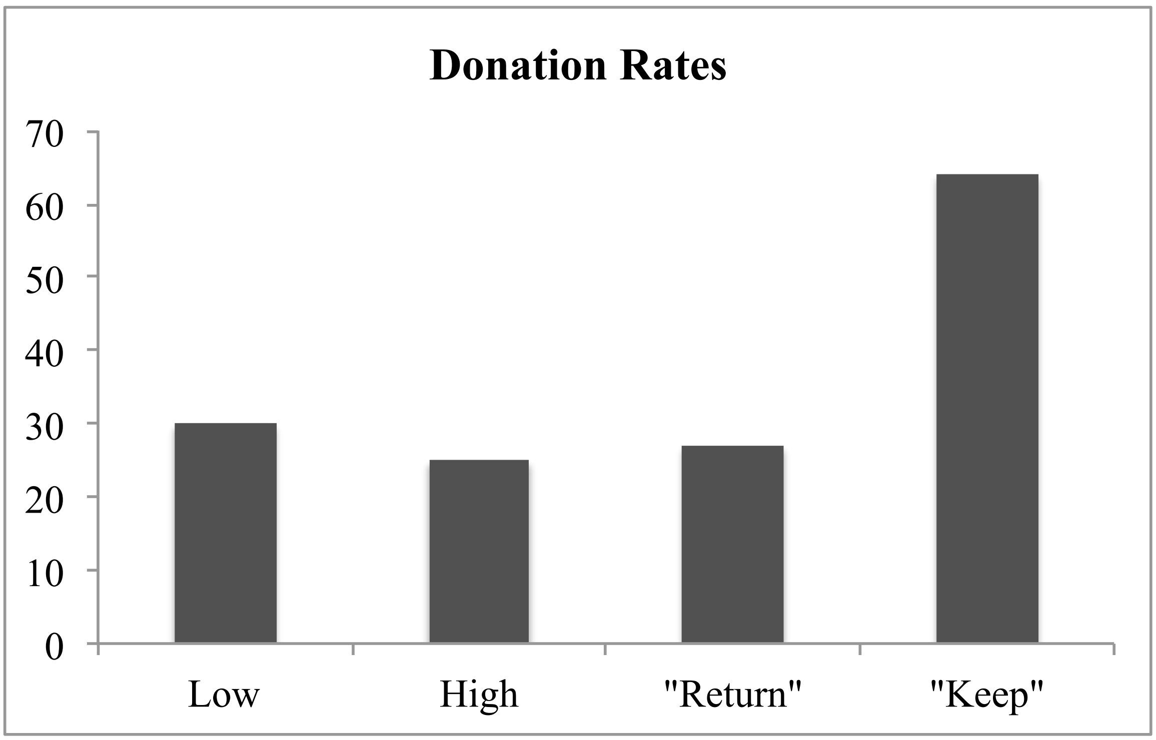


Figure 4.2: Fraction of Participants Who Donated by Treatment

4.8 Tables

Table 4.1: Results by Treatment

Treatment	**Option**	**Sender($)**	**Receiver($)**	***N***	**Lying**	**Delay**	**Informed**
Incentive	X	20	10	57	65%	No	No
	Y	10	20				
Incentive Delay	X	20	10	36	75%	Yes	No
	Y	10	20				
Incentive Choice	X	20	10	38	61%	Yes	Yes
	Y	10	20				
Informed Incentive	X	20	15	62	63%	No	Yes
	Y	15	20				
Baseline	X	20	15	63	48%	-	-
	Y	15	20				
No Incentive	X	10	10	54	24%	No	No
	Y	20	20				
Incentive Reverse	X	20	10	48	60%	No	No
	Y	10	20				

Table 4.2: Payoffs Used by Treatment

Treatment	**Payment Promised**	**Money Given by Mistake**	**Donation ($)**	N
Low	10	-	1	40
High	14	-	1	40
Mistake	10	4	1	80

References

Akerlof, G. A. and R. E. Kranton (2000): "Economics and Identity," *Quarterly Journal of Economics*, 115, 715–753.

Ali, M. M. (1977): "Probability and Utility Estimates for Racetrack Bettors," *Journal of Political Economy*, 85, 803–815.

Amodio, D. M., P. G. Devine, and E. Harmon-Jones (2007): "A dynamic model of guilt: implications for motivation and self-regulation in the context of prejudice." *Psychological science*, 18, 524–30.

Anderson, L. R., F. J. DiTraglia, and J. R. Gerlach (2011): "Measuring altruism in a public goods experiment: acomparison of U.S. and Czech subjects," *Experimental Economics*, 14, 426–437.

Andrade, E. B. and D. Ariely (2009): "The enduring impact of transient emotions on decision making," *Organizational Behavior and Human Decision Processes*, 109, 1–8.

Andrade, E. B. and T. H. Ho (2009): "Gaming Emotions in Social Interactions," *Journal of Consumer Research*, 36, 539–552.

Andrade, E. B. and G. Iyer (2009): "Planned Versus Actual Betting in Sequential Gambles," *Journal of Marketing Research*, 46, 372–383.

Andreoni, J. (1988): "Privately Provided Public Goods In A Large Economy - The Limits Of Altruism," *Journal of Public Economics*, 35, 57–73.

——— (1989): "Giving with Impure Altruism: Applications to charity and ricardian equivalence," *The Journal of Political Economy*, 97, 1447–1458.

——— (1990): "Impure altruism and donations to public goods: a theory of warm-glow giving," *The Economic Journal*, 100, 464–477.

——— (1993): "An experimental test of the public-goods crowding-out hypothesis," *American Economic Review*, 83, 1317–1327.

Andreoni, J., J. M. Rao, and H. Trachtman (2011): "Avoiding The Ask : A Field Experiment on Altruism , Empathy , and Charitable Giving," *mimeo.*

Ariely, D., A. Bracha, and S. Meier (2009): "Doing Good or Doing Well? Image Motivation and Monetary Incentives in Behaving Prosocially," *American Economic Review*, 99, 544–555.

Barber, B. M., Y.-T. Lee, Y.-J. Liu, and T. Odean (2006): "Just How Much Do Individual Investors Lose by Trading?" *Review of Financial Studies*, 22, 609–632.

Barberis, N. (2013): "Thirty Years of Prospect Theory in Economics: A Review and Assessment," *Journal of Economic Perspectives*, 27, 173–196.

Barberis, N., M. Huang, and T. Santos (2001): "Prospect Theory and Asset Prices," *The Quarterly Journal of Economics*, 116, 1–53.

Barberis, N., M. Huang, and R. H. Thaler (2006): "Individual Preferences, Monetary Gambles, and Stock Market Participation: A Case for Narrow Framing," *American Economic Review*, 96, 1069–1090.

Barberis, N. and W. Xiong (2012): "Realization Utility," *Journal of Financial Economics*, 104, 251–271.

Barkan, R. and J. R. Busemeyer (1999): "Changing Plans Dynamic Inconsistency and the Effect of Experience on the Reference Point," *Psychonomic Bulletin & Review*, 6, 547–554.

Battigalli, P. and M. Dufwenberg (2007): "Guilt in Games," *American Economic Review*, 97, 170–176.

——— (2009): "Dynamic psychological games," *Journal of Economic Theory*, 144, 1–35.

Baumeister, R. F., A. M. Stillwell, and T. F. Heatherton (1994): "Guilt: An interpersonal approach." *Psychological Bulletin*, 115, 243–267.

Becker, G. S. (1976): "A Theory of Social Interactions," *Journal of Political Economy*, 82, 1063–1093.

Benartzi, S. and R. H. Thaler (1995): "Myopic Loss Aversion and the Equity Premium Puzzle," *Quarterly Journal of Economics*, 110, 73–92.

Blount, S. (1995): "When Social Outcomes Arent Fair: The Effect of Causal Attributions on Preferences," *Organizational Behavior and Human Decision Processes*, 63, 131–144.

Blume, A. and U. Gneezy (2010): "Cognitive forward induction and coordination without common knowledge: An experimental study," *Games and Economic Behavior*, 68, 488–511.

Bordalo, P., N. Gennaioli, and A. Shleifer (2012): "Salience Theory of Choice Under Risk," *The Quarterly Journal of Economics*, 127, 1243–1285.

Bosman, R., M. Sutter, and F. van Winden (2005): "The impact of real effort and emotions in the power-to-take game," *Journal of Economic Psychology*, 26, 407–429.

Bracha, A. and D. J. Brown (2012): "Affective decision making: A theory of optimism bias," *Games and Economic Behavior*, 75, 67–80.

Camerer, C. F. and G. Loewenstein (2004): "Behavioral Economics: Past, Present, Future," in *Advances in Behavioral Economics*, ed. by C. F. Camerer, G. Loewenstein, and M. Rabin, New York: Princeton University Press, chap. 1, 3–51.

Caplin, A. (2003): "Fear as a Policy Instrument," in *Time and Decision: Economic and Psychological Perspectives on Intertemporal Choice*, ed. by G. F. Loewenstein, D. Read, and R. F. Baumeister, New York, NY: Russell Sage Foundation, July, 441–458.

Caplin, A. and J. Leahy (2001): "Psychological Expected Utility Theory and Anticipatory Feelings," *Quarterly Journal of Economics*, 116, 55–79.

——— (2004): "The Supply of Information by a Concerned Expert," *The Economic Journal*, 114, 487–505.

Cappelen, A. W., A. D. Hole, E. O. Sø rensen, and B. Tungodden (2007): "The Pluralism of Fairness Ideals: An Experimental Approach," *American Economic Review*, 97, 818–827.

Cappelen, A. W., J. Konow, E. O. Sø rensen, and B. Tungodden (2013): "Just Luck: An Experimental Study of Risk-Taking and Fairness," *American Economic Review*, 103, 1398–1413.

Card, D. and G. B. Dahl (2011): "Family Violence and Football: The Effect of Unexpected Emotional Cues on Violent Behavior," *The Quarterly Journal of Economics*, 126, 103–143.

Charness, G. (2004): "Attribution and Reciprocity in an Experimental Labor Market," *Journal of Labor Economics*, 22, 665–688.

Charness, G. and M. Dufwenberg (2006): "Promises and Partnership," *Econometrica*, 74, 1579–1601.

Charness, G. and U. Gneezy (2010): "Portfolio Choice and Risk Attitudes: An Experiment," *Economic Inquiry*, 48, 133–146.

——— (2012): "Strong Evidence for Gender Differences in Risk Taking," *Journal of Economic Behavior & Organization*, 83, 50–58.

Charness, G. and M. Rabin (2002): "Understanding social preferences with simple tests," *The Quarterly Journal of Economics*, 117, 817–869.

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

Dana, J., R. A. Weber, and J. X. Kuang (2007): "Exploiting moral wiggle room: experiments demonstrating an illusory preference for fairness," *Economic Theory*, 33, 67–80.

Davis, P. a., T. Woodman, and N. Callow (2010): "Better out than in: The influence of anger regulation on physical performance," *Personality and Individual Differences*, 49, 457–460.

de Hooge, I. (2012): "The exemplary social emotion guilt: Not so relationship-oriented when another person repairs for you," *Cognition & Emotion*, 7, 1189–1207.

de Hooge, I. E., M. Zeelenberg, and S. M. Breugelmans (2007): "Moral sentiments and cooperation: Differential influences of shame and guilt," *Cognition & Emotion*, 21, 1025–1042.

Deci, E. L. (1971): "Effects of externally mediated rewards on intrinsic motivation." *Journal of Personality and Social Psychology*, 18, 105–115.

Della Vigna, S. and U. Malmendier (2006): "Paying Not to Go to the Gym," *American Economic Review*, 96, 694–719.

DellaVigna, S., J. A. List, and U. Malmendier (2012): "Testing for Altruism and Social Pressure in Charitable Giving," *The Quarterly Journal of Economics*, 127, 1–56.

Dillenberger, D. and K. Rozen (2014): "History-Dependent Risk Attitude," *mimeo.*

Ditraglia, F. J. (2006): "Experimental Public Goods in Prague and Williamsburg: An International Comparison," *mimeo, College of William and Mary.*

Dreber, A., D. G. Rand, D. Fudenberg, and M. A. Nowak (2008): "Winners don't punish." *Nature*, 452, 348–51.

Dufwenberg, M. and U. Gneezy (2000): "Measuring Beliefs in an Experimental Lost Wallet Game," *Games and Economic Behavior*, 30, 163–182.

Dunn, E. W., L. B. Aknin, and M. I. Norton (2008): "Spending money on others promotes happiness." *Science (New York, N.Y.)*, 319, 1687–8.

Elster, J. (1998): "Emotions and Economic Theory," *Journal of Economic Literature*, 36, 47–74.

Fehr, E. and S. Gächter (2002): "Altruistic punishment in humans." *Nature*, 415, 137–40.

Frydman, C., N. Barberis, C. Camerer, P. Bossaerts, and A. Rangel (2012): "Using Neural Data to Test a Theory of Investor Behavior: An Application to Realization Utility," *Journal of Finance, forthcoming.*

Fudenberg, D., D. G. Rand, and A. Dreber (2012): "Slow to Anger and Fast to Forgive: Cooperation in an Uncertain World," *American Economic Review*, 102, 720–749.

Fuster, A. and S. Meier (2009): "Another Hidden Cost of Incentives: The Detrimental Effect on Norm Enforcement," *Management Science*, 56, 57–70.

Gibson, D. E., M. E. Schweitzer, R. R. Callister, and B. Gray (2009): "The Influence of Anger Expressions on Outcomes in Organizations," *Negotiations and Conflict Management Research*, 2, 236–262.

Gino, F. and M. Schweitzer (2008): "Blinded by anger or feeling the love: how emotions influence advice taking." *Journal of Applied Psychology*, 93, 1165–1173.

Gneezy, A., A. Imas, A. Brown, L. D. Nelson, and M. I. Norton (2011a): "Paying to Be Nice: Consistency and Costly Prosocial Behavior," *Management Science*, 58, 179–187.

Gneezy, U. (2005): "Deception: The role of consequences," *The American Economic Review*, 95, 384–394.

Gneezy, U., A. Imas, and K. Madarász (2012): "Conscience Accounting : Emotional Dynamics and Social Behavior," *TE Series 563, London School of Economics.*

——— (2014): "Conscience Accounting : Emotional Dynamics and Social Behavior," *Management Science.*

Gneezy, U. and J. A. List (2006): "Putting Behavioral Economics to Work: Testing for Gift Exchange in Labor Markets Using Field Experiments," *Econometrica*, 74, 1365–1384.

Gneezy, U., S. Meier, and P. Rey-Biel (2011b): "When and Why Incentives (Don't) Work to Modify Behavior," *Journal of Economic Perspectives*, 25, 191–210.

Gneezy, U. and J. Potters (1997): "An Experiment on Risk Taking and Evaluation Periods," *The Quarterly Journal of Economics*, 112, 631–645.

Gneezy, U. and A. Rustichini (2000): "Pay Enough or Don't Pay at All," *Quarterly Journal of Economics*, 115, 791–810.

Haigh, M. S. and J. A. List (2005): "Do Professional Traders Exhibit Myopic Loss Aversion? An Experimental Analysis," *The Journal of Finance*, 60, 523–534.

Heath, C. (1995): "Escalation and De-escalation of Commitment in Response to Sunk Costs: The Role of Budgeting in Mental Accounting," *Organizational Behavior and Human Decision Processes*, 62, 38–54.

Heyman, J. and D. Ariely (2004): "Effort for payment. A tale of two markets." *Psychological science*, 15, 787–93.

Hopfensitz, A. and E. Reuben (2009): "The importance of emotions for the effectiveness of social punishment," *The Economic Journal*, 119, 1534–1559.

Hsee, C. K. and Y. Rottenstreich (2004): "Music, pandas, and muggers: on the affective psychology of value." *Journal of experimental psychology. General*, 133, 23–30.

Ingersoll, J. E. and L. J. Jin (2013): "Realization Utility with Reference-Dependent Preferences," *Review of Financial Studies*, 26, 723–767.

Kahneman, D., J. L. Knetsch, and R. H. Thaler (1986): "Fairness and the assumptions of economics," *The Journal of Business*, 59, S285.

Kahneman, D. and D. Lovallo (1993): "Timid Choices and Bold Forecasts: A Cognitive Perspective on Risk Taking," *Management Science*, 39, 17–31.

Kahneman, D. and A. Tversky (1979): "Prospect Theory: An Analysis of Decision under Risk," *Econometrica*, 47, 263–292.

Kaustia, M. and S. Knupfer (2008): "Do Investors Overweight Personal Experience? Evidence from IPO Subscriptions," *The Journal of Finance*, 63, 2679–2702.

Keltner, D. and J. S. Lerner (2010): "Emotion," in *The Handbook of Social Psychology*, ed. by D. T. Gilbert, S. Fiske, and G. Lindsay, New York, NY: McGraw Hill, 312–347.

Khan, U. and R. Dhar (2006): "Licensing Effect in Consumer Choice," XLIII, 259–266.

Kircher, P., S. Ludwig, and A. Sandroni (2011): "On the difference between social and private goods," *mimeo*.

Knapp, a. and M. S. Clark (1991): "Some Detrimental Effects of Negative Mood on Individuals' Ability to Solve Resource Dilemmas," *Personality and Social Psychology Bulletin*, 17, 678–688.

Koszegi, B. (2006): "Emotional Agency," *Quarterly Journal of Economics*, 121, 121–155.

Koszegi, B. and M. Rabin (2006): "A Model of Reference-Dependent Preferences," *The Quarterly Journal of Economics*, 121, 1133–1165.

——— (2009): "Reference-Dependent Consumption Plans," *American Economic Review*, 99, 909–936.

Krupka, E. L. and R. A. Weber (2013): "IDENTIFYING SOCIAL NORMS USING COORDINATION GAMES: WHY DOES DICTATOR GAME SHARING VARY?" *Journal of the European Economic Association*, 11, 495–524.

Kullgren, J. T., A. B. Troxel, G. Loewenstein, D. A. Asch, L. A. Norton, L. Wesby, Y. Tao, J. Zhu, and K. G. Volpp (2013): "Individual-versus group-based financial incentives for weight loss: a randomized, controlled trial." *Annals of Internal Medicine*, 158, 505–14.

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

Linardi, S. and M. A. McConnell (2011): "No excuses for good behavior: Volunteering and the social environment," *Journal of Public Economics*, 95, 445–454.

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

Loewenstein, G. (1987): "Anticipation and the Valuation of Delayed Consumption," *Economic Journal*, 97, 666–684.

——— (1996): "Out of Control : Visceral Influences on Behavior," *Organizational Behavior and Human Decision Processes*, 65, 272–292.

——— (2000): "Emotions in Economic Theory and Economic Behavior," *American Economic Review*, 90, 426–432.

Loewenstein, G., T. O'Donoghue, and M. Rabin (2003): "Projection Bias in Predicting Future Utility," *The Quarterly Journal of Economics*, 118, 1209–1248.

Loewenstein, G. F., L. Thompson, and M. H. Bazerman (1989): "Social utility and decision making in interpersonal contexts." *Journal of Personality and Social Psychology*, 57, 426–441.

Loewenstein, G. F., E. U. Weber, C. K. Hsee, and N. Welch (2001): "Risk as Feelings." *Psychological Bulletin*, 127, 267–286.

Meier, S. (2007): "Do Subsidies Increase Charitable Giving in the Long Run? Matching Donations in a Field Experiment," *Journal of the European Economic Association*, 5, 1203–1222.

Monin, B. and D. T. Miller (2001): "Moral Credentials and the Expression of Prejudice," *Journal of Personality and Social Psychology*, 81, 33–43.

Neilson, W. S. (2009): "A theory of kindness, reluctance, and shame for social preferences," *Games and Economic Behavior*, 66, 394–403.

Norton, M. I., L. Anik, L. B. Aknin, E. W. Dunn, and J. Quoidbach (2012): "Prosocial Incentives Increase Employee Satisfaction and Team Performance," *mimeo*, 1–18.

Odean, T. (1998): "Are Investors Reluctant to Realize Their Losses?" *The Journal of Finance*, LIII, 1775–1798.

O'Donoghue, T. and M. Rabin (1999): "Doing It Now or Later Doing It Now or Later," *American Economic Review*, 89, 103–124.

Offerman, T. (2002): "Hurting hurts more than helping helps," *European Economic Review*, 46, 1423–1437.

Pagel, M. (2012): "Expectations-Based Reference-Dependent Life-Cycle Consumption," *mimeo*.

Pillutla, M. M. and J. K. Murnighan (1996): "Unfairness, anger, and spite: Emotional rejections of ultimatum offers," *Organizational Behavior and Human Decision Processes*, 68, 208–224.

Rabin, M. (2000): "Risk Aversion and Expected-Utility Theory: A Calibration Theorem," *Econometrica*, 68, 1281–1292.

Rabin, M. and G. Weizsäcker (2009): "Narrow Bracketing and Dominated Choices," *American Economic Review*, 99, 1508–1543.

Read, D., G. Loewenstein, and M. Rabin (1999): "Choice Bracketing," *Journal of Risk and Uncertainty*, 19, 171–197.

Rick, S. and G. Loewenstein (2008): "Hypermotivation," *Journal of Marketing Research*, 45, 645–648.

Sachdeva, S., R. Iliev, and D. L. Medin (2009): "Sinning saints and saintly sinners: the paradox of moral self-regulation." *Psychological science*, 20, 523–8.

Savage, L. J. (1954): *The Foundations of Statistics*, New York: Wiley.

Schmid, F. and M. Trede (1996): "Testing for first-order stochastic dominance : a new distribution-free test," *Journal of the Royal Statistical Society. Series D (The Statistician)*, 45, 371–380.

Shefrin, H. and M. Statman (1985): "The Disposition to Sell Winners Too Early and Ride Losers Too Long: Theory and Evidence," *Journal of Finance*, 40, 777–790.

Shiv, B., G. Loewenstein, and A. Bechara (2006): "The Dark Side of Emotion in Decision-Making: When Individuals with Decreased Emotional Reactions Make More Advantageous Decisions." *Cognitive Brain Research*, 23, 85–92.

Shiv, B., G. Loewenstein, A. Bechara, H. Damasio, and A. R. Damasio (2005): "Investment behavior and the negative side of emotion." *Psychological science*, 16, 435–9.

Small, D. A. and G. Loewenstein (2003): "Helping a Victim or Helping the Victim : Altruism and Identifiability," *The Journal of Risk and Uncertainty*, 26, 5–16.

Small, D. a., G. Loewenstein, and P. Slovic (2007): "Sympathy and callousness: The impact of deliberative thought on donations to identifiable and statistical victims," *Organizational Behavior and Human Decision Processes*, 102, 143–153.

Smith, G., M. Levere, and R. Kurtzman (2009): "Poker Player Behavior After Big Wins and Big Losses," *Management Science*, 55, 1547–1555.

Sutter, M. (2009): "Deception Through Telling the Truth?! Experimental Evidence From Individuals and Teams*," *The Economic Journal*, 119, 47–60.

Sutter, M., S. Haigner, and M. G. Kocher (2010): "Choosing the Carrot or the Stick? Endogenous Institutional Choice in Social Dilemma Situations," *Review of Economic Studies*, 77, 1540–1566.

Tamir, M., C. Mitchell, and J. J. Gross (2008): "Hedonic and instrumental motives in anger regulation." *Psychological science*, 19, 324–8.

Thaler, R. (1985): "Mental Accounting and Consumer Choice," *Marketing science*, 4, 199–214.

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

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

Tiedens, L. Z. and S. Linton (2001): "Judgment under emotional certainty and uncertainty: the effects of specific emotions on information processing." *Journal of personality and social psychology*, 81, 973–88.

Tonin, M. and M. Vlassopoulos (2010): "Disentangling the sources of pro-socially motivated effort: A field experiment," *Journal of Public Economics*, 94, 1086–1092.

——— (2012): "Social incentives matter: Evidence from an online real effort experiment," *IZA Discussion Paper No. 6716.*

Van Boven, L., K. White, and M. Huber (2009): "Immediacy bias in emotion perception: current emotions seem more intense than previous emotions." *Journal of experimental psychology. General*, 138, 368–82.

van Kleef, G. A., C. K. W. De Dreu, and A. S. R. Manstead (2004): "The Interpersonal Effects of Anger and Happiness on Negotiaion Behavior and Outcomes," *Journal of Personality and Social Psychology*, 86, 57–76.

van Winden, F., M. Krawczyk, and A. Hopfensitz (2011): "Investment, resolution of risk, and the role of affect," *Journal of Economic Psychology*, 32, 918–939.

Vesterlund, L. (2003): "The informational value of sequential fundraising," *Journal of Public Economics*, 87, 627–657.

Wakker, P. and A. Tversky (1993): "An Axiomatization of Cumulative Prospect Theory," *Journal of Risk and Uncertainty*, 176, 147–175.

Weber, M. and C. F. Camerer (1998): "The Disposition Effect in Securities Trading: An Experimental Analysis," *Journal of Economic Behavior & Organization*, 33, 167–184.

Weber, M. and H. Zuchel (2005): "How Do Prior Outcomes Affect Risky Choice? Further Evidence on the House-Money Effect and Escalation of Commitment," *Decision Analysis*, 2, 30–43.

Xiao, E. and D. Houser (2005): "Emotion expression in human punishment behavior." *Proceedings of the National Academy of Sciences of the United States of America*, 102, 7398–401.

Zhong, C.-B., G. Ku, R. B. Lount, and J. K. Murnighan (2009): "Compensatory Ethics," *Journal of Business Ethics*, 92, 323–339.