Game Shows, Gambles, and Economic Behavior

ISBN: 978 90 316 0401 2

Cover illustration: ©iStock.com/TSchon
Cover design: Crasborn Graphic Designers bno, Valkenburg a.d. Geul

This book is no. **589** of the Tinbergen Institute Research Series, established through cooperation between Thela Thesis and the Tinbergen Institute. A list of books which already appeared in the series can be found in the back.

Game Shows, Gambles, and Economic Behavior

Spelshows, kansspelen en economisch gedrag

Proefschrift

ter verkrijging van de graad van doctor aan de
Erasmus Universiteit Rotterdam
op gezag van de rector magnificus

Prof.dr. H.A.P. Pols

en volgens besluit van het College voor Promoties.

De openbare verdediging zal plaatsvinden op

donderdag 25 september 2014 om 15:30 uur

door

Dennie van Dolder

geboren te Zeist

ERASMUS UNIVERSITEIT ROTTERDAM

Promotiecommissie

Promotoren:	Prof.dr. H. Bleichrodt
	Prof.dr. P.P. Wakker
Overige leden:	Prof.dr. C.N. Noussair
	Prof.dr. S. Gächter
	Dr. A. Baillon
Copromotor:	Dr. M.J. van den Assem

Acknowledgements

There are many people who have supported and encouraged me over the years. Hereby I want to take the opportunity to thank those who have done so.

I want to start off by thanking my supervisors, Peter Wakker and Han Bleichrodt, and my co-supervisor Martijn van den Assem.

Peter and Han gave me maximum freedom and support to pursue my own projects. At the weekly Thursday meetings, Peter sharpens the mind of his students by means of group discussions. The education provided by Peter, both in these meetings and in day-to-day interactions, has shaped my thinking about the field in a profound way. Han always has an open door to discuss both research and practical matters, and I cannot thank him enough for all the help provided over the past years. Looking back over the years the only regret one can have is that we have not collaborated on more projects.

When I had just started my PhD in Rotterdam, Martijn approached me with the idea to work on a game show paper. This is now the second chapter of this thesis. Martijn is creative, open-minded, and an excellent lecturer. I have learned a lot from him in terms of communicating research in an accessible manner. Our collaboration has been very fruitful and at the moment it is difficult to keep track of all our joint projects.

Peter and Han have built an excellent behavioral economics research group in Rotterdam and I am grateful for the opportunity to interact with the many bright people working there. Rogier Potter van Loon was the best office mate a person could wish for and Jan Stoop was the best office crasher a person could wish for. The conversations between us three were some of the nicest moments in Rotterdam. Aurélien Baillon functioned as a daily supervisor and I am grateful to have been able to profit from his advice and wisdom over the years. If I had to point to one person as being an academic role model, it would most likely be Aurélien. Further thanks goes to Kirsten Rohde, Arthur Attema, Julia Müller, Amit Kothiyal, Vitalie Spinu, Umut Keskin, Asli Selim, Zhenxing Huang, Zhihua Li, Chen Li, Yu Gao, Ning Liu, Tong Wang, Uyanga Thurmunkh, and Ilke Aydogan for the positive interactions over the years.

Further thanks goes to Richard Thaler for inviting me to spend time at the University of Chicago Booth School of Business. I am also grateful to Mike Yeomans, who proved to be an excellent guide in Chicago. George Wu welcomed me in his lab group meetings and allowed me to present research there on multiple occasions. I am grateful for this opportunity, for the feedback received there, and the people I was able to meet as a result. Next I want to thank everybody who helped make my stay at Chicago so amazing. I will avoid writing an entire list as this will surely mean that I will forget many people who have been amazing during my time there.

Thanks to all co-authors on chapters of this thesis. In addition to the people already mentioned above these are Mohammed Abdellaoui, Guido Baltussen, Colin Camerer, and Olivier l'Haridon. Further thanks to Simon Gächter and Charles Noussair for their judgment of this thesis.

Financial support from the Netherlands Organization for Scientific Research (VICI NWO Grant 453-06-001) and the Tinbergen Institute is gratefully acknowledged.

I could not have made it to a PhD without the human capital investments made prior to coming to Rotterdam. Undoubtedly, the five years I spend at the sociology department at Utrecht University were the most formative years of my life. The quality of the education at this department is world class and the choice to start my studies there was probably the best choice in my life thus far. I want to thank Vincent Buskens, Rense Corten, Henk Flap, Werner Raub, Tanja van der Lippe, Frank van Tubergen, Beate Volker, Jeroen Weesie, and all other lecturers and fellow students who made Utrecht such a great experience. After completing my research master in Sociology and Social Reserach in Utrecht, I went to Nottingham to study behavioral economics. I thank Robin Cubitt, Simon Gächter, Martin Sefton, and Chris Starmer for the excellent field courses on behavioral and experimental economics, and all other lecturers and fellow students I encountered there. Furthermore, I like to thank all those great individuals whom I have had the pleasure to interact with at summer schools and conferences throughout the years.

Finally, a warm thanks to all friends and family who have supported me over the years. A special thanks goes out to my parents, who have always been there for me and have encouraged me to pursue my own interest and dreams.

Contents

List of Tables

List of Figures

Chapter 1 | Introduction

Over the past decades, laboratory experiments have become an important source of data within economics. For most of the twentieth century, the majority of economists held the view that it is impossible to use experiments in order to test the predictions made by economic theory (Friedman, 1953; Lipsey, 1979; Samuelson and Nordhaus, 1985). Although seminal economic experiments were conducted throughout the last centuries, for a considerable period these studies remained largely isolated pieces of work (Bardsley *et al.*, 2010). This is no longer the case today. Experimental papers are published frequently even in the most general top journals and many economic departments around the world have computer laboratories suited to perform the most complex forms of experiments.

Despite this strong increase in experimental work, many economists remain skeptical about the value of laboratory experiments. The main concern remains that of external validity: the degree to which findings from experiments can be generalized to other environments. Binmore (1999), for example, argues that the conditions in the typical experiment are such that economic theory cannot reasonably be expected to work well. In particular, he suggests that economic theory should only hold when incentives are adequately high and learning opportunities are sufficient, which is arguably not the case in many experiments. Levitt and List (2007a, 2007b, 2008) make similar arguments. They state that incentives in experiments tend to be considerably smaller than those in many naturally occurring settings, that subjects are often less familiar with the decision task in the laboratory than with those in everyday life, and that they have no opportunity to seek outside advice. They also stress that selection mechanisms in the laboratory differ from those in the field, that experiments are completed over short durations whereas real life decision are made over longer time frames, and that the nature and extent of scrutiny that subjects face in the laboratory is unparalleled in the field. Such scrutiny can be expected to induce greater levels of social behavior in the laboratory than in many field settings.

While the above arguments clearly have their merit, some nuances are in order. First, it should be stressed that (most of) the criticisms above do not uniquely apply to laboratory experiments. Binmore (1999) acknowledges that if economic theories only hold when stakes are adequately high and learning opportunities are sufficient, then we should also not expect them to hold in field settings that do not satisfy these conditions. Levitt and List (2007a) acknowledge that the arguments they put forth "apply with equal force to generalizing from data generated from naturally occurring environments" (p. 170). As a result, it is not obvious whether data from a laboratory experiment are less informative than field data to predict behavior in a different and unrelated field environment (Falk and Heckman, 2009). Second, the impact of the aforementioned factors on behavior and, therefore, the external validity of experiments are in the end empirical questions. Not just that, these questions can be well studied by employing experimental methods (Starmer, 1999). As noted by Camerer (2011), most conclusive evidence suggesting that the factors above impact behavior comes from laboratory data.

As a result, rather than disregarding laboratory data out of hand, a more constructive approach is to actively investigate whether laboratory findings are robust. Two obvious approaches to do so present themselves. First, one can try to study questions that are typically investigated in the behavioral laboratory in the field, either by conducting field experiments (Harrison and List, 2006) or by locating unique naturally occurring data that can be used for this purpose. Observing behavior under widely different conditions can be an important step in identifying which patterns are robust and which are not. Second, one can use experimental methods to map the effect of the abovementioned factors on decision-making. Such an approach can ultimately inform the construction of models that can provide a framework to transport findings between different environments or populations (Levitt and List, 2007a; Falk and Heckman, 2009). This thesis incorporates both approaches.

Chapter 2 and 3 use naturally occurring data from TV game shows in order to test behavioral hypotheses that are generally difficult to test outside of the controlled laboratory. Chapter 4 and 5 employ experiments to investigate the effect of public scrutiny on behavior. Finally Chapter 6 takes a slightly different focus. This chapter introduces a new and easily applicable method to measure utility and loss aversion, both under risk and under uncertainty and employs this method to study whether utility is the same under risk and uncertainty.

1.1 Game Shows as Natural Experiments

Since the early 90s, numerous papers have used game show data in order to test behavioral hypotheses or estimate parameters of behavioral decision models. Due to the fact that many game shows include risky and strategic decisions, it is not surprising that most of these papers study either risk preference (Gertner, 1993; Metrick, 1995; Beetsma and Schotman, 2001; Post *et al.*, 2008) or strategic behavior (Bennett and Hickman, 1993; Berk, Hughson, and Vandezande, 1996; Tenorio and Cason, 2002). More recently, game shows have also been used to study social interaction, most notably discrimination (Levitt, 2004; Antonovics, Arcidiacono and Walsh, 2005) and cooperative behavior (List, 2004, 2006; Belot *et al.*, 2010; Oberholzer-Gee *et al.*, 2010).

The reason for economists' interest in game shows is that they allow researchers to study behavior in well-designed decision problems when the stakes are large. Due to their uncontrolled nature, field data rarely allow for a clean discrimination between competing theories. Laboratory experiments do allow researchers to construct such controlled tests, but the incentives that can be offered to subjects are limited by the researcher's budget. Game show data can help breach this gap.

In addition to employing large and widely ranging stakes, game shows differ from laboratory experiments on many other dimensions as well. Participants in the laboratory tend to be volunteering students with no experience with the abstract task that they will be represented with. They have no opportunity to gain advice from friends and know that their decisions will be used with the aim of testing scientific hypotheses. Game shows differ on all these aspects. The selection mechanisms differ from show to show, but are very different from those in the laboratory. In general, there is much more diversity in demographic characteristics as participants in many shows appear to be a reasonable (middle-class) cross section of the population. Furthermore, contestants tend to be familiar with the task at hand and thus had the opportunity to prepare themselves. While they are being watched on TV, this type of scrutiny is of a different scope and nature than the scrutiny faced in laboratory experiments. From the vantage point of studying the robustness of behavioral findings outside of the laboratory, these differences have important implications. The downside is that if behavior in a game show differs from that in the laboratory, it is difficult to pinpoint which factors underlie this difference. If, however, decisions from the laboratory are replicated in the very different environment of a TV game show, this provides a strong signal regarding the robustness of these findings.

In Chapter 2, we study cooperative behavior when large sums of money are at stake, using data from the British TV game show “Golden Balls”. At the end of each episode, contestants play a variant of the classic Prisoner’s Dilemma for large and widely ranging stakes averaging over $20,000. The variation is large: from a few dollars to about $175,000.

On average, contestants cooperate 53 percent of the time. This rate is similar to that observed in prior experimental work (Dawes and Thaler, 1988; Sally, 1995). With respect to demographic characteristics, we observe that young males are less cooperative than young females and that this differences changes with age. Older men are more cooperative than younger men, while such an age effect is absent for women. As a result, from about 46 years onwards men in this game show are more likely to cooperate than women.

The dynamic nature of the show allows us to test a number of interesting behavioral hypotheses. In support of the claim that people have reciprocal preferences, we find that contestants are significantly less likely to cooperate if their opponent in the final has previously attempted to vote them of the show. Furthermore, we find that cheap talk is predictive of behavior. Making a promise is the strongest predictor of behavior in the show: those who make a promise are 31 percentage points more likely to cooperate. In spite of the predictive power of promises, contestants do not condition their choices on the promises made by their opponent. More generally, we find that contestants do not appear to condition their choice on any factor that predicts the cooperation likelihood of their opponent. This implies that contestants either lack the ability or ignore the possibility to reliably interpret information about the expected behavior of others, or do not have a preference for matching the other's choice. Given our finding that people do reciprocate votes against them, the former explanation seems more likely.

Finally, our results provide support for the view that attitudes are strongly influenced by context. We find unusually high rates of cooperation when the luck of the game reduces the stakes to "merely" a few hundred Pounds. Such amounts are tiny in the light of the thousands and even tens of thousands the game is often played for, but would be considered very large in any laboratory setting. Supporting the view that contestants evaluate money amounts in relative rather than absolute terms, we find that in the early days of the show, when the contestants have not had an opportunity to watch the show on TV and are still learning what kind of stakes are to be expected, cooperation rates appear to be influenced by the salient but normatively irrelevant value representing the sum they could have been playing for with a lucky selection of balls. In particular, the higher this maximum jackpot, and thus the smaller the actual jackpot appears, the greater the likelihood that contestants cooperate. Across episodes, this effect vanishes, suggesting that expectations about the stakes become well informed.

Chapter 3 investigates bargaining behavior using data from the British TV game show "Divided". In Divided, three contestants collectively build up a jackpot through answering general quiz questions. Across episodes, their jackpot ultimately ranges from approximately $10,000 to $185,000, and averages over $50,000. In the second phase of the game, the team's accumulated money amount is divided into three unequal parts of, for example, 60, 30 and 10 percent. Contestants in turn have to claim one of these shares.

If they do not immediately agree on who takes which share, they have 100 seconds to negotiate and reach consensus. With each second they take they lose one percentage point of the initial jackpot, and after 100 seconds there is nothing left. This final stage can thus be seen as a natural bargaining experiment where the "subjects" have to unanimously decide on the allocation of three indivisible shares, in a format that allows face-to-face communication and incorporates (close to) continuous costs to bargaining.

Overall, 50 percent of the jackpot value is lost in bargaining. Because the jackpot is determined by teams' answers to trivia questions, we are able to investigate the influence of entitlements on bargaining. We find that equity concerns play an important role in the bargaining process. Contestants that contributed more to the communal jackpot claim a larger share, are less likely to lower their claim during the bargaining process, and end up with a larger fraction of the jackpot.

Contestants making hardball announcements, by adding a statement to their initial claim that they will not back down from it, act accordingly. These contestants are less likely to back down from their initial claim. As a result, they increase their likelihood of taking the top share home. Due to the increased bargaining costs, however, this strategy does not increase their earnings in an absolute sense and lowers the earnings of their opponents. There is no evidence of a first-mover advantage: the order in which contestants get to make their claims does not influence the claims they stake, nor the outcomes reached. Finally, there is little evidence that behavior and outcomes are related to demographic characteristics.

1.2 Public Scrutiny

Over recent years, the nature and degree of scrutiny that laboratory subjects face has been a major point of criticism of experiments in economics (Levitt and List, 2007a). Interestingly enough, scrutiny has thus far predominantly been considered as a disturbing factor in tasks in which morality and wealth are competing objectives (Levitt and List, 2007a, 2007b). The two chapters investigate whether public scrutiny also influences behavior in economic tasks that do not incorporate a moral component.

Chapter 4 investigates the impact of public scrutiny on risky choice. It presents the results from two incentivized experiments that mimic the game of the TV show Deal or No Deal. Both experiments include laboratory and limelight treatments. In the laboratory treatments, subjects make their decisions anonymously under conditions typically employed in economic experiments. In the limelight treatments, subjects make their choices in a simulated game show environment, including a live audience, game show host, and video cameras.

Comparing behavior between the laboratory and limelight treatments, we find that subjects are more risk averse in the limelight than in the anonymity of a typical behavioral laboratory. Estimates of structural choice models indicate that the impact of the limelight on risk preference parameters is substantial. At the same time, however, subjects in both treatments show path dependent behavior; they take more risk if the game develops either substantially worse or substantially better than expected. As a result, under both conditions our simple prospect-theory inspired model with a path-dependent reference point outperforms expected utility of wealth in explaining subjects' choices.

In addition, three other findings emerge. First, exploiting a design difference between the sets of experiments we find that ambiguity aversion depends on being in the limelight. Although we find substantial evidence for ambiguity aversion in the limelight, we do not observe it in the laboratory. Second, passive experience gained by watching others play the game does not affects loss aversion in particular or risk aversion in general. Finally, estimates from all treatments suggest that preferences are based on imperfectly updated expectations.

Chapter 5 investigates the effect of social cues on behavior in both individual choice and interaction tasks. Recent literature has shown that even subtle cues of being watched can influence behavior in tasks have a moral component. The presence of pictures of a pair of eyes, or an eye-like stimulus, in an otherwise anonymous experimental setting leads to increased donations to strangers (Haley and Fessler, 2005; Rigdon *et al.*, 2009; Oda *et al.*, 2011; Nettle *et al.*, 2013), increased donations to public goods (Burnham and Hare, 2007), and induces greater disapproval of moral transgressions (Bourrat *et al.*, 2011). The fact that relatively subtle social cues can influence behavior is significant for the experimentalist as it suggests that, even in anonymous laboratory settings, pro-social behavior cannot be viewed as being purely intrinsic (Haley and Fessler, 2005; Jaeggi *et al.*, 2010).

Thus far, the effect of pictures of eyes has only been investigated in tasks that include a moral component. The fact that actual public scrutiny also influences behavior in individual choice tasks that have no moral component begs the question whether pictures of eyes will also affect behavior in such tasks. Furthermore, it remains unclear whether the effect of eyes is something special or whether pictures of eyes constitute one among many social cues that produce the same effect.

Chapter 5 presents the results of an internet experiment designed to acquire a better understanding of the effect of pictures of eyes on human behavior. First, in order to investigate whether the effect of eyes is limited to interaction tasks, we expand the range of tasks to include individual choice tasks that have no moral component. Second, in order to investigate whether different social cues have similar effects, we compare the effect of pictures of eyes with a different condition in which we present subjects with pictures of other students (peers).

Our results suggest that the effect of pictures of eyes is limited to interaction tasks that include a moral component and that eyes should be considered as distinct from other social cues, such as reminders of peers. Whereas pictures of eyes uniformly enhanced pro-social behavior in our experiment, this is not the case for reminders of peers. Furthermore, reminders of peers trigger more rational behavior in individual choice tasks that have no moral component, whereas pictures of eyes do not affect behavior in such tasks.

1.3 Measuring Utility

Chapter 6 introduces a new and easily applicable method to measure utility and loss aversion, both under risk and under uncertainty. This method extends the trade-off method of Wakker and Deneffe (1996) by allowing standard sequences (sequences of outcomes for which the utility differences between successive elements is constant) to pass through the reference point. Thus, we make the trade-off method robust to sign dependence and allow for standard sequences that include gains, losses, and the reference point. As with the traditional trade-off method, our method requires no simplifying assumptions about utility or event weighting.

We employ our method to test whether the utility function has the same shape under risk and uncertainty. This test is critical for models that capture ambiguity aversion through a difference in event weighting between risk and uncertainty, such as multiple priors and prospect theory. We cannot reject the hypothesis that utility and loss aversion are the same for risk and uncertainty, suggesting that utility primarily reflects attitudes towards outcomes. Under both risk and uncertainty, we find S-shaped utility (concave for gains, convex for losses) and substantial loss aversion.

Chapter 2 | Split or Steal? Cooperative Behavior when the Stakes are Large

This chapter examines cooperative behavior when large sums of money are at stake, using data from the TV game show "Golden Balls". At the end of each episode, contestants play a variant on the classic Prisoner's Dilemma for large and widely ranging stakes averaging over $20,000. Cooperation is surprisingly high for amounts that would normally be considered consequential but look tiny in their current context, what we call a "big peanuts" phenomenon. Utilizing the prior interaction among contestants, we find evidence that people have reciprocal preferences. Surprisingly, there is little support for conditional cooperation in our sample. That is, players do not seem to be more likely to cooperate if their opponent might be expected to cooperate. Further, we replicate earlier findings that males are less cooperative than females, but this gender effect reverses for older contestants because men become increasingly cooperative as their age increases.

This chapter is based on the paper "Split or Steal? Cooperative Behavior when the Stakes Are Large", co-authored by Martijn J. van den Assem and Richard H. Thaler, and published in Management Science (van den Assem, van Dolder and Thaler, 2012). We are grateful to Guido Baltussen, Han Bleichrodt, Uri Gneezy for their many constructive and valuable comments. We thank Endemol UK and in particular Tara Ali and Tom Blakeson for providing us with information and recordings of "Golden Balls", and Shirley Kremer, Thomas Meijer and Sanne van Bemmel for their skillful research assistance. The chapter has benefited from discussions with seminar participants at the Erasmus University of Rotterdam and Tilburg University, and with participants of the Decision & Uncertainty Workshop 2010 at HEC Paris, BDRM 2010 Pittsburgh, SABE 2010 San Diego, and SPUDM 2011 Kingston upon Thames.

2.1 Introduction

Cooperation is vital for the functioning of society, and the organizations and communities that form its fabric. Not surprisingly, cooperative behavior is the focus of many studies across a wide range of scientific disciplines, including psychology (Dawes, 1980; Dawes and Messick, 2000), sociology (Marwell and Ames, 1979, 1980; Raub and Snijders, 1997), economics (Ledyard, 1995; Fehr and Gächter, 2000a; Fischbacher and Gächter, 2010), political science (Ostrom, Walker and Gardner, 1992) and biology (Gardner and West, 2004; West, Griffin and Gardner, 2007). The key question in this literature is why humans cooperate even in situations in which doing so is not in line with their material self-interest.

While cooperation is ubiquitous in social life and an important topic for all kinds of economic interaction, field data rarely allow for a clean discrimination among competing theories. Because carefully designed laboratory experiments do allow for such rigorous comparisons, laboratory experiments have provided numerous important insights into cooperative behavior, and the resulting rich literature forms the basis of most of our knowledge on human cooperation. Still, laboratory settings inevitably have limitations that some argue may hinder the generalization of findings to situations beyond the context of the lab (Levitt and List, 2007a, 2008). Subjects are often volunteering students who thus constitute a non-random sample of the population at large. Also, they generally have less familiarity with decision tasks in the laboratory than with those in everyday life, no opportunity to seek advice from friends or experts, and they know that their behavior is examined in detail.

From an economic perspective, another obvious drawback to lab studies is that the financial stakes employed tend to be relatively small. Even those experiments that utilize relatively large payoffs do not involve amounts in excess of a few hundred dollars (*e.g.,* Hoffman, McCabe and Smith, 1996a; List and Cherry, 2000; Carpenter, Verhoogen and Burks, 2005), giving rise to the question to what extent findings will generalize to situations of significant economic importance. One solution is to perform experiments in low-income countries, where small nominal amounts carry a larger value. In the domain of social interaction, such experiments are, for example, employed by Slonim and Roth (1998), Cameron (1999), Fehr, Fischbacher and Tougareva (2002), Munier and Zaharia (2002), Johansson-Stenman, Mahmud and Martinsson (2005), and Kocher, Martinsson and Visser (2008). While this might appear an ideal approach, it has its own drawbacks. Culture, for example, has been shown to play an important role in social interaction (Henrich *et al.,* 2001, 2004; Herrmann, Thöni and Gächter, 2008), making it difficult to generalize findings from low-income countries.[1] And while the stakes in these experiments are larger than commonly employed, they still rarely exceed a few months' wages.

In the current chapter, we study cooperative behavior using another source of data, namely the behavior of contestants on the British TV game show "Golden Balls". Although the game show setting is an unusual environment, it has the benefit of employing large and varying stakes. Furthermore, game shows are markedly different from laboratory experiments in terms of participant selection, scrutiny, and familiarity of participants with the decision task. Combined with the strict and well-defined rules, game shows can therefore provide unique opportunities to investigate the robustness of existing laboratory findings.

Because game shows are often competitive in nature and ask contestants to make risky or strategic choices, is it not surprising that they have mostly been used to study decision making under risk (*e.g.,* Gertner, 1993; Metrick, 1995; Beetsma and Schotman, 2001; Post *et al.,* 2008) or strategic reasoning (*e.g.,* Bennett and Hickman, 1993; Berk, Hughson and Vandezande, 1996; Tenorio and Cason, 2002). More recently, however, game shows have also been used to study social interaction, in particular discrimination (Levitt, 2004; Antonovics, Arcidiacono and Walsh, 2005) and cooperative behavior (List, 2004a, 2006; Belot, Bhaskar and van de Ven, 2010a; Oberholzer-Gee, Waldfogel and White, 2010). The current chapter is in the latter category.

[1] Interestingly, though not generally acknowledged, this argument at the same time questions the universal applicability of the many findings from higher-income countries, including ours. We refer to Henrich, Heine and Norenzayan (2010) for a discussion on this issue.

In the final stage of "Golden Balls", contestants make a choice on whether or not to cooperate in a variant of the famous Prisoner's Dilemma. In particular, the two final contestants independently have to decide whether they want to "split" or "steal" the jackpot. If both contestants choose "split", they share the jackpot equally. If one chooses "split" and the other chooses "steal", the one who steals takes the jackpot and the other gets nothing. If they both "steal", both go home empty-handed. On average, the jackpot is over $20,000. The variation is large: from a few dollars to about $175,000.

If we assume that each player only cares about maximizing her immediate financial payoff, the choice problem in "Golden Balls" can be labeled as a "weak" form of the Prisoner's Dilemma.[2] Where in the classic form of the Prisoner's Dilemma defecting strictly dominates cooperating, here defecting only weakly dominates cooperating: choosing "steal" always does at least as well, and sometimes better than choosing "split". Of course, contestants may consider other factors aside from their own monetary payoff when deciding which choice to make. Much experimental research suggests that people have social preferences in the sense that the payoffs to others enter their utility functions. For discussions, see, for example, Fehr and Gächter (1998, 2000b), Fehr and Schmidt (1999), Bolton and Ockenfels (2000), Charness and Rabin (2002), Camerer (2003), Fehr and Gintis (2007), and Cooper and Kagel (2009).

The fact that the show is aired on TV of course creates a set of rather special circumstances that could affect our results, although there is little existing theory to suggest what the effect of a large TV audience would be. One might argue that players would not want to be seen as a "jerk" on national television and so would be more likely to cooperate, but one can also argue that a player would not want to been seen as a "sucker" (or someone who cannot detect the weakly dominant solution to a simple game) in public.[3] The public nature of the choice could also magnify subtle features created by

[2] Rapoport (1988) introduced this terminology. For the sake of brevity, we will simply use the term Prisoner's Dilemma to refer to the game studied here.

[3] Most studies related to the issue of observability indicate that people display more other-regarding behavior when they are or feel more subject to public scrutiny (see, for example, Hoffman, McCabe and Smith, 1996b; Rege and Telle, 2004; Haley and Fessler, 2005), but there is also contradictory evidence

the fact that the game is a "weak" form of the Prisoner's Dilemma. Specifically, if a player thinks that the other player will steal, she might decide to split on the grounds that it costs her nothing to appear "nice" on TV. These complications do not render our results uninteresting, but do need to be incorporated in any attempt to evaluate how our results should be interpreted in the context of existing theories and experimental findings on cooperation.

Although "Golden Balls" is unique in its format, the show shares the Prisoner's Dilemma element with a few game shows from other countries including "Friend or Foe" (US, aired in 2002-2003) and "*Deelt ie 't of deelt ie 't niet*?" (in English: "Will he share or not?"; Netherlands, 2002). These two shows have been studied in four different papers. List (2004a, 2006) and Oberholzer-Gee, Waldfogel and White (2010) analyze data from "Friend or Foe", and Belot, Bhaskar and van de Ven (2010a) use the Dutch show. List focuses on the effects of demographic variables such as gender, race and age. Studying the same game show, Oberholzer-Gee, Waldfogel and White compare the behavior in the first season of the show with later seasons in which the contestants have had a chance to observe prior episodes. Finally, Belot, Bhaskar and van de Ven find that making a promise to cooperate prior to the decision is positively related to cooperation if the promise was voluntary, but not if the host has elicited it.

In this chapter we replicate many of the earlier investigations, but also undertake several novel analyses that are possible due to some unique features in the format of "Golden Balls". The way the stakes are determined and the very wide range they cover provide the basis for new insights into the effect of stakes and context. The dynamic setting of the show enables us to look at reciprocity in cooperative behavior, and also at the effect of earlier deceitful behavior.

In our sample, individual players on average cooperate 53 percent of the time. Although this rate is similar to earlier findings from the experimental literature (Dawes and Thaler,

(Dufwenberg and Muren, 2006). Kerr (1999) suggests that the effect depends on conditions related to social expectations and sanctions.

1988; Sally, 1995), direct comparisons are hampered by systematic differences in the stakes, the visibility of decisions, characteristics of the subjects, and preceding opportunities for communication or other social interaction.

We find only limited support for the notion that cooperation will decrease if the stakes get significant. The cooperation rate is unusually high when the stakes lie in the low range of our sample, perhaps because contestants think that for so little money (relatively speaking) they might as well cooperate in public. Cooperation does decline with the stakes for stakes below the median, but plateaus at around 45 percent for medium to large amounts.

The high cooperation rate for relatively small stakes suggests that context can convert money amounts that would normally be considered consequential or "big" into amounts that are perceived to be small, just "peanuts". This idea is supported by our finding that cooperation is not only based on the actual stakes but also on what the jackpot potentially could have been. This effect is especially pronounced for those who appeared in the earlier episodes of the show and had no or little opportunity to watch the show on TV and learn what sizes are large or small in the context of this game.

A special property of "Golden Balls" is the interaction that occurs among contestants prior to the final. Utilizing the dynamic setting, we find evidence that contestants show some tendency toward reciprocity. Among contestants whose final opponent has attempted to vote them off the show, the propensity to cooperate is significantly lower. Contestants do not appear to reciprocate against opponents who have lied earlier in the game. Lying seems to be accepted here, similar to bluffing in poker. A possible reason for this is that, in contrast to a vote cast against someone, lying is a defensive act that is not aimed at anyone in particular.

Surprisingly, we find little evidence that contestants' propensity to cooperate depends positively on the likelihood that their opponent will cooperate. While an opponent's promise to cooperate is a strong predictor of her actual choice, contestants appear not to be more likely to cooperate if their opponent might be expected to cooperate. Our final main result is that young males cooperate less than young females. This difference decreases and even reverses as age increases and men become increasingly cooperative.

The chapter proceeds as follows. Section 2.2 describes the game show in more detail, discusses our data and presents descriptive statistics. Sections 2.3 – 2.7 cover the various possible factors behind cooperative behavior included in our analysis. Each of these sections provides related literature and other background, explains the variables that we use, and discusses the results of our Probit regression analyses. Section 2.8 concludes.

2.2 Game Show and Data

Description of "Golden Balls"

The TV game show "Golden Balls" was developed by the Dutch production company Endemol. It debuted on the ITV network in the United Kingdom in June 2007 and ran until December 2009. Each episode consists of four rounds and starts with four contestants, usually two men and two women.

In Round 1, twelve golden balls are randomly drawn from the "Golden Bank", a lottery machine containing one hundred "golden" balls. Each of these balls has a hidden cash amount inside, ranging from a minimum of £10 to a maximum of £75,000.[4] Contestants know that this is the range for the amounts in the balls, but they do not know the precise distribution (though this becomes clearer over time as the show is aired). At a later stage of the game, a subset of the cash balls drawn will contribute to the final jackpot. Also, four balls hiding the word "killer" inside are mixed with the twelve cash balls. Killer balls are undesirable in a way we will explain below. From the sixteen balls, each contestant receives four balls at random. For each contestant, two are placed on the front row with their contents - either a cash amount or the word "killer" - openly displayed; the other two are placed on the back row and their contents are known by the particular contestant alone. (Poker players can think of this as two "up" cards and two "down" cards.) The

[4] Values in British pounds can be translated into US dollars using a rate of $1,75 per pound, an approximate average of the exchange rate during the period in which the show ran.

contestants now have to decide by vote which player will be kicked off the show. Because the balls of voted-off contestants are removed from the game and the remaining contestants' balls matter for the ultimate jackpot, there is a strong incentive to retain players with high value balls and kick off players with low value balls or killer balls.

Before the voting starts, each contestant publicly announces the contents of the balls on her back row (knowing that these values will subsequently be revealed, but only after the vote). Then, the four contestants together have an open discussion in which they can voice their evaluation of other players' statements and their opinion of who should be voted off. Each player then anonymously casts a vote against one specific opponent. After the votes are tallied, the player who received the most votes leaves the game.[5] Lastly, all the players reveal the values of their back row balls, and differences between the actual values and the previous claims are noted.

In Round 2, two additional cash balls from the lottery machine and one extra killer ball are added to the twelve remaining balls from Round 1. The fifteen balls are then randomly allocated to the three contestants. Each of them receives two balls on her front row and three on her (hidden) back row. Similar to Round 1, contestants make (cheap talk) statements on the balls on their back row, a round of banter follows, votes are cast anonymously and tallied, the player who receives the most votes leaves the game, and all hidden ball values are revealed.[6] Two players and their ten balls proceed to Round 3.

Round 3 determines the size of the final jackpot. First, one additional killer ball is mixed with the ten balls from Round 2. Then five of the balls are selected sequentially at random. If a ball selected is a cash ball, its face value is added to the jackpot. If a killer ball is drawn, the current cumulative jackpot is divided by ten. For example, if the first two

[5] If two contestants receive two votes each, their opponents openly discuss who they want to keep in the show. If they cannot decide, a decision is made at random. If all four contestants receive one vote each, contestants openly attempt to form a coalition against one specific contestant. Again, if they cannot decide, a decision is made at random.

[6] The procedure in the case of a tie is similar to that in Round 1. Tie-breaking occurs by discussion or by random draw if no agreement is reached.

balls were £50,000 and £1,000 and the third is a killer ball, the level of the jackpot is reduced from £51,000 to £5,100. A killer ball does not affect the jackpot contribution of cash balls drawn thereafter. If the fourth and fifth ball in our example were another killer ball and £25,000, respectively, then the actual jackpot would be £25,510. Note that this round is a completely stochastic process, and that contestants have full information on the balls that are in play. Before the five balls are drawn, special attention is always paid to the highest possible jackpot (that is, the sum of the five largest cash values). This value, and the number of killer balls, are explicitly stressed by the game show host.

After Round 3 determined the jackpot, the contestants play a variant of the Prisoner's Dilemma in the fourth and final round. Each contestant receives two golden balls. One of the balls says "split" and the other says "steal" on the inside. The contestants then simultaneously have to decide which ball they want to play. If both choose "split", they share the jackpot equally. If one chooses "split" and the other chooses "steal", the contestant who steals takes the whole jackpot and the other gets nothing. If they both choose "steal", both go home empty-handed. Before each contestant makes her actual decision, a brief time period is reserved for a discussion between the players in which they can make non-binding promises, ask about intentions, or attempt to get assurances of cooperative behavior. This is the final round of cheap talk. Importantly, the contestants have not met before the game starts, and have no opportunity before or during the show to make any kind of collusive agreement.

A relevant question is how the contestants are selected. A spokeswoman of Endemol informed us that anyone can apply to be on "Golden Balls" by submitting a detailed application form. Shortlisted contestants are then invited to an audition in order to determine their skills at playing the game, their character and their suitability to appear on a TV show such as Golden Balls. Producers watch tapings of these auditions and put together shows such that, according to the producers, "a good mix of characters" is represented on each show. Thus while the contestants are not a random sample of society, the selection process does not seem to create any obvious confounds with the analyses we conduct here.

Data and Descriptive Game Characteristics

We examine the “split” and “steal” decisions of 574 final contestants appearing in 287 episodes aired between June 2007 (when the show was introduced) and December 2009. During this period, 288 episodes were aired, and, at the time of writing, no further episodes were aired thereafter. Recordings from the show and additional information such as recording and airing dates were kindly provided by Endemol’s local production company Endemol UK. The one missing episode could not be supplied because it was not present in their archives.[7]

For each episode we collected data on the relevant observables in the show, such as the hidden and visible ball values, statements made by contestants, the votes, the jackpot size, and the decision to “split” or “steal” at the end. Some variables were estimated based on contestants’ physical appearance and on information provided in the introductory talk and other conversations during the show.

Table 2.1 displays some descriptive characteristics of the game. Cash balls drawn from the lottery machine during the first two rounds have a mean value of £5,654 and a median of £1,500. Clearly, the distribution is positively skewed. The mean value of the cash balls taken to Round 3 is £6,775, which is statistically significantly greater than the average value of all cash balls in the show, implying that the contestants are successful in using their votes to keep high-value balls in play and eliminate small ones. The average number of killer balls in the game at the start of Round 3 is 3.14, significantly less than the 3.67 we would statistically expect if voting was random. Contestants thus also seem successful in eliminating killer balls from the game. Unreported analyses of contestants’ voting behavior clearly show that contestants indeed try to vote off the opponents that have the worst set of balls on their front row.

[7] Sixteen episodes in our data set feature returning contestants. In twelve of these, players who previously had lost in the final (opponents “stole” while they themselves chose to “split”) get a second chance. In four, unlucky players who had been voted off in the first game round receive a second chance. We do not find that returning contestants behave differently, and, unless stated otherwise hereafter, excluding them from our analyses does not materially affect our results.

At the start of Round 3, special attention is paid to the highest possible jackpot. Dependent on the cash balls and killer balls taken to this stage, this maximum varies between £5,000 and £168,100, with a mean of £51,493 and a median of £41,150. The actual jackpot for which contestants play the Prisoner's Dilemma game is generally considerably smaller due to the skewed distribution of cash ball values and the effect of killer balls, but still has a mean size of £13,416 and a median of £4,300. These amounts are many times the amounts typically used in laboratory experiments, and also large sums relative to the median gross weekly earnings of £397 in the UK in April 2009 (Office for National Statistics 2009). About half of the time, the jackpot in our show exceeds three months of median UK earnings, and 21 percent of the contestants decide over a jackpot that is even larger than a median annual salary (the third quartile in our sample is at £18,350). The stakes are also large compared to the two other game shows employed in earlier analyses of cooperative behavior: in "Friend or Foe", the average is about \$3,500 (List, 2004a, 2006; Oberholzer-Gee, Waldfogel and White, 2010), and for the Dutch show, Belot, Bhaskar and van de Ven (2010a) report a median of €1,683. The wide range of the jackpot in our sample is caused by its random construction, by the highly skewed distribution of cash ball values, and by the effect of killer balls. The largest jackpot was played for in an exhilarating episode from March 2008; trainee accountant Sarah stole the entire jackpot of £100,150 from collection agent Stephen.[8]

[8] A video clip of this episode is widely available on the Internet, for example through YouTube.

Table 2.1: Selected Game Show Characteristics

The table shows selected characteristics for the British TV game show “Golden Balls”, extracted from our sample of 287 episodes. *Cash ball (overall)* is the monetary value of a cash ball drawn from the lottery machine in the first or second round of the game. *Cash ball (Round 3)* is the monetary value of a cash ball that is in play at the start of the third round. *No. of killer balls (Round 3)* describes the number of killer balls that are in play at the start of the third round. *Potential jackpot (Round 3)* is the jackpot size that is attained during the third round if the best-case scenario would occur. *Jackpot* describes the actual size of the jackpot. *Decision* is a contestant’s decision in the Prisoner’s Dilemma at the end of the show, with a value of 1 for “split” and 0 for “steal”. *Prize won (if non-zero)* records the take home prize for a contestant who made it to the final (if she did not leave empty-handed). All monetary values are in UK Pounds (£1.00 ≈ $1.75).

	N	Mean	Stdev	Min	Median	Max
Cash ball (overall)	4,018	5,653.92	10,478.49	10.00	1,500.00	75,000.00
Cash ball (Round 3)	2,257	6,775.15	12,204.39	10.00	1,600.00	75,000.00
No. of killer balls (Round 3)	287	3.14	0.90	1.00	3.00	6.00
Potential jackpot (Round 3)	287	51,493.08	31,386.69	5,000.00	41,150.00	168,100.00
Jackpot	287	13,416.09	19,182.98	2.85	4,300.00	100,150.00
Decision (split=1)	574	0.53	0.50	0.00	1.00	1.00
Prize won	574	4,850.55	11,821.06	0.00	38.75	100,150.00
Prize won if non-zero	303	9,188.82	15,004.52	1.83	2,175.00	100,150.00

For the jackpot to be awarded, at least one player needs to cooperate. We find that 52.8 percent of the contestants decide to “split”. While this might seem high, the rate is actually remarkably similar to earlier experimental evidence (see, for example, Sally, 1995). In our sample, both players split the jackpot 31 percent of the time, one splits while the other one steals occurs in 44 percent of the shows, and in the remaining 25 percent of the shows both players steal. The efficiency rate in terms of the percentage of jackpots that is actually awarded thus amounts to 75 percent. The efficiency rate obtained by dividing the sum of earnings across all episodes by the sum of all jackpots is slightly lower at 72 percent. (The difference in efficiency results from contestants’ lower propensity to cooperate when the stakes are larger; we explore this effect in detail later.) These simple statistics are a first indication that contestants do not condition their behavior on that of their opponent. Given the average cooperation rate, we would expect to observe (split, split) in 28 percent of the cases and (steal, steal) 22 percent of the time if the individual decisions in our sample were randomly matched. Although the actual percentages are higher (31 and 25), the differences are relatively small considering that each pair of contestants operates under highly similar conditions (same jackpot, same potential jackpot, and many shared unobserved conditions).

On average, a finalist goes home with £4,851, but the median prize is only £39 because 47 percent of the contestants get nothing. The 303 contestants who end up with a non-zero prize take home £9,189 on average, with a median of £2,175. It is worth noting that would we have run this show as an experiment ourselves, the total costs in subject payoffs alone would have been £2.8 million.

Modeling the Decision to "Split" or "Steal"

In the following sections, we will analyze the decisions to "split" or "steal" the jackpot using a binary Probit model. We assume that when people enter the final round they have a latent propensity to "split" y^*, where $y^* \in (-\infty, \infty)$. Furthermore, we assume that this latent propensity is a linear function of personal demographic characteristics x and context characteristics z, in the form $y^* = x'\beta + z'\gamma + u$, where β and γ are parameter vectors and u represents an unobserved stochastic component. We do not observe the latent propensity to "split" directly, but only the actual decision y, where $y = 1$ if a contestant chooses "split" and $y = 0$ if a contestant chooses "steal". We impose the observation criterion $y = \mathbf{1}(y^* > 0)$, where $\mathbf{1}(.)$ is the indicator function taking the value of 1 if $y^* > 0$ and 0 otherwise. Assuming that the stochastic component has a standard normal distribution, or $u \sim N(0,1)$, leads to the binary Probit model of the form $\Pr(y = 1 | x, z) = \Phi(x'\beta + z'\gamma)$, where $\Phi(.)$ is the standard normal cumulative distribution function. Using this framework we estimate the parameter vectors β and γ using maximum likelihood estimation. We allow for the possibility that the decisions of contestants within the same episode are correlated by performing a clustering correction on the standard errors (see, for example, Wooldridge, 2003).

Because coefficients in a Probit model do not have an immediate intuitive economic meaning due to the inherent nonlinearities, we will follow the common approach of reporting marginal effects instead. More specifically, the marginal effects we report apply to the medians of the explanatory variables, with two exceptions. To highlight some of the interaction effects that we find in our data, we set Age to be 20 and Transmissions (the number of times the show has aired at the time of recording) to be 0. The resulting “representative agent” is a 20 year old white female without higher education, who lives in a relatively small town and plays the final of our game for a jackpot of £4,300, which potentially could have been £41,150. For dummy variables we consider the effect of a discrete change from 0 to 1. As noted by Ai and Norton (2003), the traditional way of calculating marginal effects and their standard errors is not valid for interaction terms, and we therefore apply the alternative method they propose. For the sake of consistency, we report significance levels that apply to the marginal effects, though these levels do not differ materially from the significance levels for the original regression coefficients. Original coefficients and their significance levels are available from the authors upon request.

2.3 Demographic Characteristics

First, we investigate how various demographic characteristics are related to the propensity to cooperate. Our later analyses include these demographic variables as control variables.

In previous studies examining the relations between demographic characteristics and cooperative behavior, most attention has been directed to gender. Psychologists have a long history when it comes to investigating the relation between gender and behavior, and, over the past decade, economists have become increasingly interested in gender effects as well. The standard finding is that women act more pro-socially than males, but the reverse is also found.[9] For contextual settings similar to ours, List (2004a, 2006), Oberholzer-Gee, Waldfogel and White (2010) and Belot, Bhaskar and van de Ven (2010a)

[9] One possible cause for the varying results is that males and females respond differently to specific contextual settings of the experiments (Croson and Gneezy, 2009). For example, if women are more risk averse than men, this may lead to different social behavior in situations in which risk is involved (Eckel and Grossman, 2008).

report that women are more cooperative than men, although some results are only marginally significant.

For other demographic characteristics, the experimental findings are also mixed. Carpenter, Daniere and Takahashi (2004), for example, run public good experiments with symbolic but costly punishment in Bangkok and Ho Chi Minh City. They find that in Bangkok males and higher educated subjects contribute more, while there is no significant age effect. The same experiment in Ho Chi Minh City, however, shows the opposite findings: males and higher educated subjects cooperate less and age increases cooperation. Gächter, Herrmann and Thöni (2004) find no influence of background characteristics in a one-shot public good experiment with Russian subjects.

In order to add to this literature we will explore the effect of various demographic characteristics on cooperative behavior. We employ the following set of variables:

- "Gender" is a dummy variable indicating whether a contestant is male (1) or female (0).
- "Age" is a continuous variable measuring the contestant's age in years. In many instances the contestant's age is not explicitly mentioned during the show. In these cases we estimate age on the basis of physical appearance and other helpful information such as the age of children.
- "Race" is a dummy variable indicating whether a contestant is white (1) or non-white (0). We apply such a broad distinction because the large majority of contestants are white.
- "City" and "London" are two dummy variables that are constructed in order to distinguish contestants that live in major urban areas from those that reside in more rural surroundings. Contestants' city or county of residence is always an integral part of the introductory talk. City indicates whether a contestant lives in a large urban area (1) or not (0). We define a large urban area as a conurbation with a population

exceeding 250,000 inhabitants.[10] For some contestants we only know their region and not their exact town or city; we then assume a small domicile. London indicates whether a contestant lives inside (1) or outside (0) the Greater London Urban Area.

- “Education” is a dummy variable for the level of education and differentiates between those with at least a bachelor degree (1) and those without (0). Players generally do not talk about their education during the show. We therefore estimate a contestant’s level of education on the basis of her occupation, which is always explicitly mentioned when she is introduced, and on the basis of other information given in talks. Contestants who are currently enrolled in higher education and people whose job title suggests work experience equivalent to the bachelor level or higher are included in the higher education category. From the information that we have about each contestant, the proper binary values are generally clear.
- “Student” is a dummy variable indicating whether the contestant currently is a higher education (undergraduate or postgraduate) student (1) or not (0).

Estimates for Age and Education are based on the independent judgments of three research assistants, where each value is based on the assessments of two of them. When the estimates for Education were different, we decided on the most appropriate value ourselves. For Age we took the mean of the two judgments, and included our own assessment as a third input if the values of the coders diverged more than five years.

We have also attempted to collect data on contestants’ marital status and the existence of children. These topics were, however, not systematically discussed in the program and values would therefore be unknown for the large majority of our contestants. Table 2.2 summarizes all the variables that are included in our analyses, including the demographic characteristics.

[10] For England and Wales, the population data and the definitions of conurbations are taken from the UK Office for National Statistics (www.statistics.gov.uk). Similar information for Scotland, Northern Ireland and Ireland is from the General Register Office for Scotland (www.gro-scotland.gov.uk), the Northern Ireland Statistics and Research Agency (www.nisra.gov.uk) and the Central Statistics Office Ireland (www.cso.ie), respectively.

Table 2.2: Summary Statistics

The table shows descriptive statistics for the explanatory variables in our analyses of cooperative behavior based on the decisions of 574 contestants to either “split” or “steal” the jackpot in the Prisoner’s Dilemma at the end of the British TV game show “Golden Balls”. *Age* is the contestant’s age measured in years. *Gender*, *Race*, *City*, *London*, *Education* and *Student* are dummy variables taking the value of 1 if the contestant is male (*Gender*), is white (*Race*), lives in a conurbation with a population exceeding 250,000 inhabitants (*City*), is a resident of the Greater London Urban Area (*London*), has completed or is enrolled in higher education (bachelor degree or higher) or has equivalent working experience (*Education*), or is a student (*Student*), respectively. *Actual stakes* is the natural logarithm of the size of the jackpot in the Prisoner’s Dilemma game. *Potential stakes* is the natural logarithm of the highest possible jackpot at the start of the third round. *Transmissions* expresses the number of episodes that was already aired when the current episode was recorded in the studio. *Vote received from opp.* is a dummy variable taking the value of 1 if the contestant’s final opponent has tried to vote her off the program at an earlier stage of the game. *Promise* is a dummy variable taking the value of 1 if the contestant explicitly promised her opponent to “split” (or not to “steal”) the jackpot. *Lie Round 1 (Round 2)* is a dummy variable taking the value of 1 if the contestant has misrepresented her back row balls – either by overstating a cash ball or by hiding a killer ball – in the first (second) round. *Lie cash (killer) ball Round 1 (Round 2)* is a dummy variable taking the value of 1 if the contestant has overstated a cash ball (hidden a killer ball) in the first (second) round. Standard deviations for the two stakes variables and the transmissions variable are calculated across episodes (N = 287) to avoid the effect of clusters at the episode level. All monetary values are in UK Pounds (£1.00 ≈ $1.75).

	Mean	Stdev	Min	Median	Max
Demographic Characteristics					
Age	36.78	11.76	18.00	34.40	73.00
Gender (male=1)	0.47	0.50	0.00	0.00	1.00
Race (white=1)	0.92	0.27	0.00	1.00	1.00
City (large=1)	0.47	0.50	0.00	0.00	1.00
London (London=1)	0.14	0.35	0.00	0.00	1.00
Education (high=1)	0.34	0.47	0.00	0.00	1.00
Student (student=1)	0.09	0.28	0.00	0.00	1.00
Stakes and Context					
Actual stakes (log)	8.19	2.08	1.05	8.37	11.51
Potential stakes (log)	10.68	0.60	8.52	10.62	12.03
Transmissions	111.68	74.18	0.00	109.00	214.00
Reciprocal Preferences					
Vote received from opp. (yes=1)	0.05	0.22	0.00	0.00	1.00
Expectational Conditional Cooperation					
Promise (promise=1)	0.53	0.50	0.00	1.00	1.00
Past Deceitful Behavior					
Lie Round 1 (lie=1)	0.41	0.49	0.00	0.00	1.00
Lie Round 2 (lie=1)	0.36	0.48	0.00	0.00	1.00
Lie cash ball Round 1 (lie=1)	0.24	0.42	0.00	0.00	1.00
Lie cash ball Round 2 (lie=1)	0.15	0.36	0.00	0.00	1.00
Lie killer ball Round 1 (lie=1)	0.21	0.41	0.00	0.00	1.00
Lie killer ball Round 2 (lie=1)	0.24	0.43	0.00	0.00	1.00

Table 2.3: Binary Probit Regression Results [1/2]

The table displays results from the Probit regression analyses of contestants' decisions to "split" (1) or "steal" (0) the jackpot in the Prisoner's Dilemma at the end of the British TV game show "Golden Balls". *First (Second) half* is a dummy variable taking the value of 1 if less (more) than 50 percent of the episodes in our sample were already aired when the current episode was recorded in the studio. Definitions of other variables are as in Table 2.2. For each explanatory variable, the marginal effect is shown for a representative agent who takes the median value on all variables, except for *Age* and *Transmissions*, which are set to 20 and 0, respectively. Standard errors are corrected for clustering at the episode level, *p*-values are in parentheses.

	Model 2.1		Model 2.2		Model 2.3		Model 2.4		Model 2.5	
Demographic Characteristics										
Age	0.002	(0.422)	0.002	(0.345)	0.003	(0.311)	0.002	(0.352)	0.002	(0.403)
Gender (male=1)	-0.221	(0.002)	-0.241	(0.000)	-0.233	(0.001)	-0.236	(0.001)	-0.233	(0.001)
Race (white=1)	0.134	(0.101)	0.143	(0.082)	0.139	(0.091)	0.142	(0.089)	0.149	(0.065)
City (large=1)	-0.039	(0.396)	-0.043	(0.359)	-0.045	(0.335)	-0.039	(0.405)	-0.036	(0.439)
London (London=1)	0.059	(0.402)	0.066	(0.348)	0.059	(0.400)	0.054	(0.444)	0.058	(0.406)
Education (high=1)	0.088	(0.062)	0.093	(0.053)	0.094	(0.050)	0.094	(0.049)	0.091	(0.058)
Student (student=1)	0.012	(0.888)	-0.002	(0.983)	-0.008	(0.923)	-0.013	(0.884)	-0.007	(0.933)
Age x Gender	0.011	(0.001)	0.010	(0.001)	0.010	(0.001)	0.011	(0.001)	0.010	(0.001)
Stakes and Context										
Actual stakes (log)			-0.043	(0.000)	-0.048	(0.000)	-0.049	(0.000)	-0.048	(0.000)
Potential stakes (log)					0.057	(0.139)	0.174	(0.006)		
Transmissions							-0.000	(0.722)		
Potential stakes x Transmissions							-0.001	(0.037)		
Second half (second=1)									0.030	(0.508)
Potential stakes x First half									0.139	(0.005)
Potential stakes x Second half									-0.035	(0.543)
Wald chi^2 (df)	34.87(8)		51.61(9)		52.57(10)		57.87(12)		62.35(12)	
Log pseudo-likelihood	-379.78		-371.55		-370.42		-368.29		-367.51	
McFadden R^2	0.043		0.064		0.067		0.072		0.074	
N	574		574		574		574		574	
Number of clusters	287		287		287		287		287	

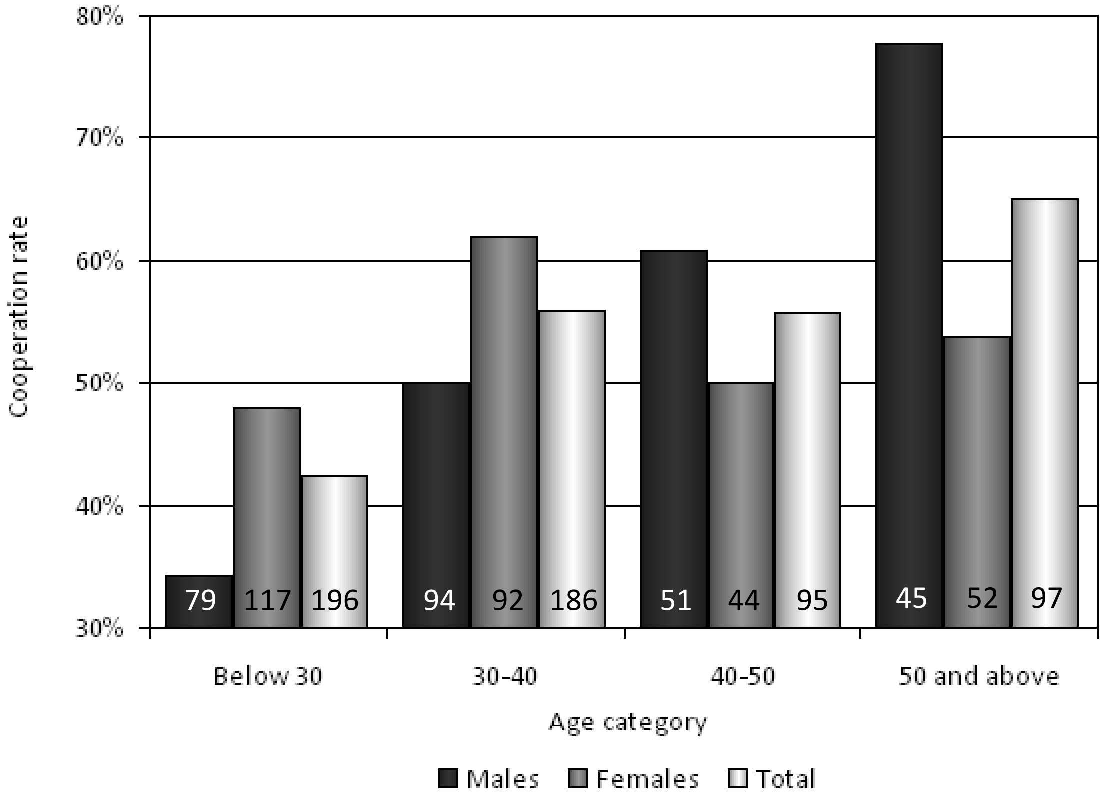


Figure 2.1: Age and the Propensity to Cooperate for Males and Females

The figure displays the relative frequency of contestants who decide to “split” across various age intervals. Bars depict the percentage of cooperators within specific age brackets for males, females and the aggregate, respectively. For each category, the number of contestants is displayed at the bottom of the bar.

Table 2.3, Model 2.1 shows the regression results for a model that includes demographic characteristics only. To be able to distinguish both general gender and age effects as well as a possible interaction effect, the interaction of gender and age is also included. The results show that, relative to our representative 20 year-old female agent, young males are 22 percentage points less likely to cooperate (p = 0.002). In line with past results by List (2004a) and Carpenter, Connolly and Myers (2008), this difference disappears when age increases. The effect of age is significantly different for males and females (p = 0.001). Women do not become significantly more or less cooperative when age increases (p = 0.422). Men, on the other hand, do have a higher propensity to cooperate as they are older: their cooperation rate increases by more than one percentage point per year (p = 0.000; untabulated).[11] Contrary to the two previous studies, we find that the gender difference not only disappears as age increases, but actually reverses; males become significantly more likely to "split" from age 46 onwards. Figure 2.1 displays observed cooperation rates at different age levels for males, females and aggregates, clearly depicting an age effect for men. Further analyses show that there is no evidence of a quadratic age effect, neither for men nor for women. We have also experimented with specifications where the (semi-) continuous age variable is replaced by a set of dummy variables that represent various age groups. The results are economically and statistically similar.

When it comes to race we find weak evidence that whites are more likely (about 13 percentage points) to cooperate than non-whites (p = 0.101; $p < 0.10$ in the models discussed hereafter). List (2004a, 2006) and Oberholzer-Gee, Waldfogel and White (2010) report a similar pattern, yet in more conventional experiments the reverse is often found (see, for example, Cox, Lobel and McLeod, 1991). Since possible but unobservable wealth effects could contribute to this result, it should be interpreted with caution.

Higher educated contestants are about 9 percentage points more cooperative (p = 0.062), although this effect is only marginally significant in the current model and not consistently significant across the various regression models discussed hereafter ($0.041 < p < 0.070$). Similar to the effect of race, the effect of education could be spurious due to an unobservable wealth effect.

Students are frequently used as subjects in experiments, and the reliance on such a specific subject pool is often criticized. Sears (1986), for example, extensively describes how the use of student subjects might produce misleading or mistaken conclusions about social behavior. It is therefore interesting to investigate whether there is evidence that

[11] The effect of age for males could be related to increasing dependence on others (van Lange *et al.,* 1997), or to hormonal or neurological changes as men grow older, but we are hesitant to draw conclusions in these directions for we cannot exclude that a generational effect (van Lange *et al.,* 1997; List, 2004) or a wealth effect is (partly) driving our finding.

students behave differently from others, holding other observable characteristics constant. This turns out not to be the case. Controlling for demographics such as age and education, our regression results yield no indications of a different attitude toward cooperation among students (p = 0.888).[12]

None of the residence dummy variables have a significant effect. Possibly, relatively small social differences between urban and more rural areas in the UK explain this null result.[13]

2.4 Stakes and Context

Economists typically argue that behavior will converge toward the prediction of rational self-interest if the stakes increase (*e.g.,* Rabin, 1993; Telser, 1995; Levitt and List, 2007a). The evidence from lab and field experiments is, however, not generally supportive of this view. Except for the finding that people seem to become more willing to accept relatively low offers in ultimatum bargaining games when the stakes are high, empirical research generally finds no evidence that stake size affects behavior, even when the stakes are increased up to several months' wages.[14]

Given that the stakes in "Golden Balls" are widely ranging, and, on average, considerably larger than in previous studies, the show provides an excellent opportunity to re-examine the relation between cooperation and stakes. In addition, compared to earlier game-show studies on cooperation, an advantage of "Golden Balls" is that the stakes are mainly built up by a random process and not by contestants' answers to trivia questions. The

[12] In a similar vein, van Lange *et al*. (1997) and Bellemare and Kröger (2007) do not detect a difference, whereas Carpenter, Connolly and Myers (2008) and Egas and Riedl (2008) do report a negative bias.

[13] In experiments conducted in a region of Russia where there is a large gap, Gächter and Herrmann (2011) do find that rural residents are more cooperative than urban residents.

[14] See, for example, Hoffman, McCabe and Smith (1996a), Slonim and Roth (1998), Cameron (1999), List and Cherry (2000, 2008), Fehr, Fischbacher and Tougareva (2013), Munier and Zaharia (2002), Carpenter, Verhoogen and Burks (2005), Johansson-Stenman, Mahmud and Martinsson (2005), and Kocher, Martinsson and Visser (2008). For TV game show data, List (2004, 2006) and Oberholzer-Gee, Waldfogel and White (2010) also find that cooperative behavior is practically invariant to the stakes. Belot, Bhaskar and van de Ven (2010a) report some counter-intuitive evidence that cooperation actually increases with stakes.

latter may lead to a spurious correlation because the ability to answer trivia questions may be related to unobserved background characteristics such as income, which in turn may well be related to the propensity to cooperate.

The variable that we use in our regressions is labeled “Actual stakes” and defined as the natural logarithm of the size of the jackpot.

Model 2.2 in Table 2.3 displays the regression results when the stakes are included. Clearly, cooperative behavior in our show is sensitive to the amount that is at stake. To illustrate this effect, Figure 2.2 depicts the actual and estimated cooperation rates for different stake levels. The fitted line based on our full regression model (Model 2.6 presented later on) appears to capture the pattern rather well. Cooperation is high when the stakes are relatively small: for amounts up to £500, people on average cooperate 73.4 percent of the time. The rate drops to approximately 45 percent as the stakes increase and remains relatively stable for the largest amounts. An unreported test shows that we cannot reject that the relation becomes essentially flat for stakes larger than £1,500.

While the absolute level of the stakes thus appears to have some influence on the propensity to cooperate, behavioral research suggests that people do not always evaluate prospects just in absolute terms, but rather they sometimes use relative comparisons to determine subjective values. This way, what comprises the context can strongly influence choices (Kahneman, Ritov and Schkade, 1999). In choice tasks, for example, one can increase the likelihood that a given option is chosen by adding an alternative to the choice set that is dominated by the given option but not by the other alternatives available (Huber, Payne and Puto, 1982). Also, as a consequence of the use of relative judgments, seemingly irrelevant anchors can influence how people value goods of various kinds (Green *et al.,* 1998; Ariely, Loewenstein and Prelec, 2003; Simonson and Drolet, 2004) and even (risky) monetary prospects (Johnson and Schkade, 1989).

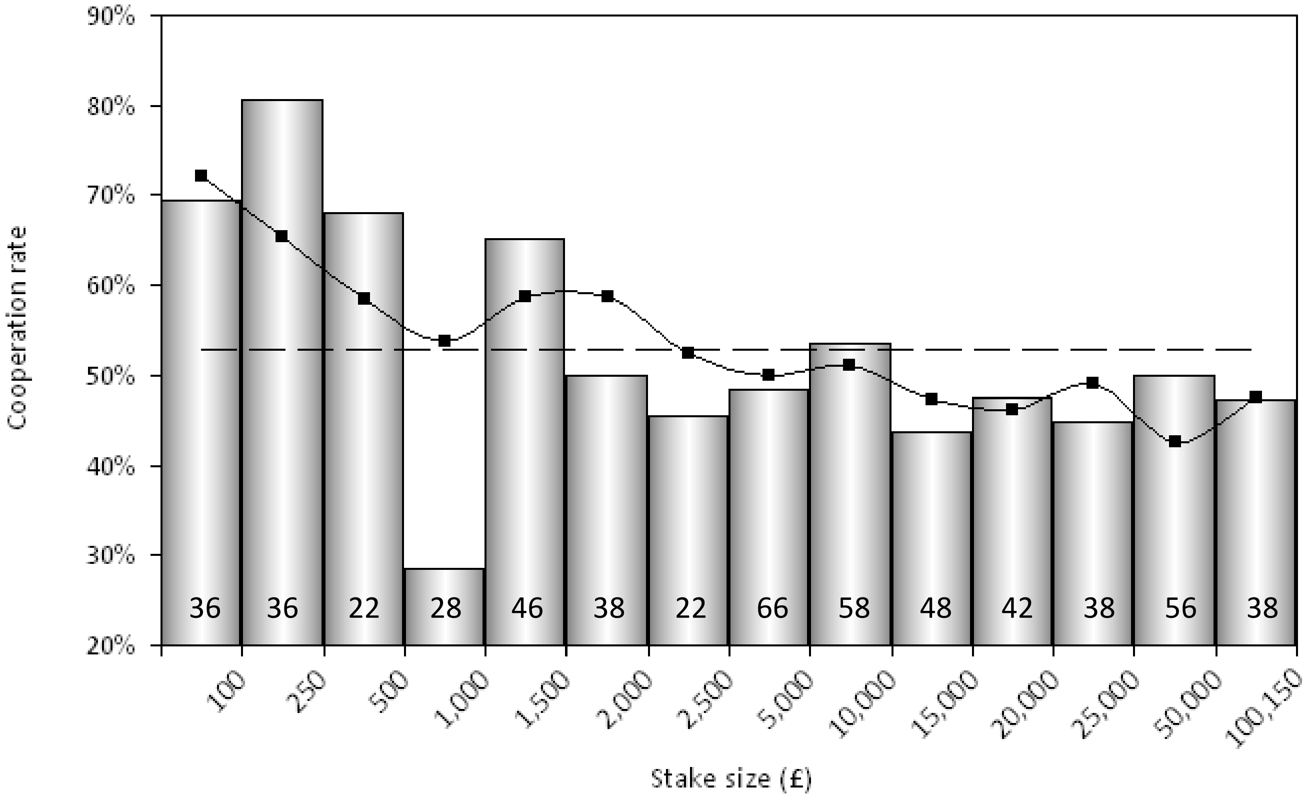


Figure 2.2: Stakes and the Propensity to Cooperate

The figure displays the relative frequency of contestants who decide to "split" across various stake intervals. Tick mark values represent the endpoints of the intervals. Each bar depicts the percentage of cooperators within a specific stake bracket. The dashed line reflects the average cooperation rate across our full sample, while the solid line connects the average estimate of the propensity to cooperate for each stake bracket. The estimates are computed using our "full model", which is Model 2.6 in Table 2.4. For each interval, the number of contestants is displayed at the bottom of the bar.

For our purposes, the question of interest is whether the game show influences the contestants' perceptions of what constitutes "serious money". Suppose that some contestants decide that for "serious" money they are willing to bear the reputational costs, if any, of defecting on national TV, but if the stakes are small, so-called "peanuts", then they will just cooperate to look good. In this scenario, we would observe the pattern of cooperation in our data: high cooperation rates for low stakes and lower cooperation for high stakes. The interesting point, however, is that the "small" stakes on this show, several hundred Pounds, are quite large relative to most experiments. Even when the contestants are playing for what seems to be peanuts, these are big peanuts indeed.

In “Deal or No Deal”, another game show that has even larger stakes than Golden Balls, Post *et al.* (2008) also find strong evidence of such a “big peanuts” phenomenon. Namely, when unlucky contestants faced decisions near the end of the show that were “merely” for thousands of euros, they displayed little or no risk aversion. In fact, some of their contestants made risk-seeking choices in such situations. The authors provide further evidence for this behavior in classroom experiments designed to mimic the show at two levels of stakes, call them “low” and “medium”. In the low stakes treatment the average prize was €40 with a maximum of €500, while in their medium stakes treatment the average prize was €400 with a maximum of €5,000. While risk aversion increased with stakes within each treatment, such an effect was not found across treatments: despite the very different money amounts, risky choices were similar for the low and medium stakes session. Choices in both conditions were even remarkably similar to those made in the actual TV show, despite the huge stakes used there (average €400,000, maximum €5,000,000).

These results suggest that a context can convert a sum of money that would normally be considered consequential into perceived peanuts. In the “Golden Balls” scenario, earlier expectations about the jackpot size or a specific value from the game might operate as an anchor or reference value by which the actual size of the jackpot is evaluated.[15] The most obvious benchmark contestants may use seems to be the maximum possible jackpot at the beginning of Round 3. Though the expected jackpot size might be an alternative candidate, it is neither salient nor easily calculated. In fact, even a rough assessment is rather complicated, particularly because of the influence of killer balls. The maximum potential jackpot, however, is always visually displayed and explicitly stressed by the game show host.

In order to test for such an effect, we include the variable “Potential stakes” in our analyses, defined as the highest possible jackpot at the start of Round 3. As with the

[15] We intentionally do not use the term “reference point” in order to avoid associations with prospect theory here, since we are hesitant to translate the elements of prospect theory to preferences in this game and to derive testable predictions.

actual stakes, we take the natural logarithm. Not surprisingly, the actual and the potential stakes variable are significantly correlated, but due to the effect of killer balls and the skewed distribution of cash balls the degree is rather limited; the Pearson correlation coefficient is $\rho = 0.30$.

Model 2.3 shows the new results. The positive sign of the potential stakes coefficient is in line with what we would expect, but the effect is statistically insignificant for the entire sample ($p = 0.139$). Interestingly, the effect becomes marginally significant if we exclude the (returning) contestants who already appeared in a previous episode of the show ($p = 0.066$; untabulated). This gives rise to the idea that the anchoring effect of the potential jackpot may decrease over time as contestants become more familiar with the show by watching it on TV. Prior shows will give contestants an impression of expected payoffs, which may help them to evaluate whether the stakes they face themselves are high or low in the context of the game and reduce the role of an episode-specific reference value such as the maximum possible jackpot size.

We explore the effect of experience by testing whether the effect of the potential jackpot changes as contestants have watched more episodes on TV. As a (noisy) proxy for how many shows a contestant has watched, we define the variable "Transmissions" as the number of different episodes broadcast on TV prior to the studio recording of the current episode. Model 2.4 in Table 2.3 displays the results. There is no significant main effect of this variable ($p = 0.722$), indicating that there is no evidence of a trend in the cooperation rate over time. However, the interaction effect of the number of transmissions and the potential stakes is significantly negative ($p = 0.037$), and implies that the anchoring effect of the maximum possible jackpot decreases by 0.10 percentage point for each previously aired episode. Controlling for this interaction effect, the effect of the maximum potential jackpot is highly significant in the early episodes ($p = 0.006$), where doubling the maximum potential jackpot increases cooperation by more than 12 percentage points.[16]

[16] Because the potential stakes and the actual stakes are correlated, the interaction of Potential stakes and Transmissions might pick up an effect of the interaction of Actual stakes and Transmissions. As a robustness

The pattern of a pronounced effect of the potential jackpot size in the earlier but not in the later shows also becomes apparent if we include dummy variables that subdivide our sample. Model 2.5, for example, uses a natural subdivision and employs a dummy variable for the first 149 episodes (0-112 transmissions prior to the recordings) and for the remaining 138 (149-214 prior transmissions). Clearly, the effect is significant across the first half our data (p = 0.005) and insignificant thereafter (p = 0.543).

2.5 Reciprocal Preferences

Reciprocity refers to a tendency to repay kindness with kindness and unkindness with unkindness. Reciprocal behavior in the field is generally embedded in long term social interaction, and reputation concerns therefore form a plausible explanation for virtually all instances where people reciprocate (Sobel, 2005). Reciprocal actions can sometimes also be explained by preferences over outcome distributions, most notably by a desire for equity or equality (Fehr and Schmidt, 1999; Bolton and Ockenfels, 2000). Interestingly, laboratory experiments in which such motivations are controlled for have convincingly demonstrated that people also have a real intrinsic preference for reciprocity.[17]

“Golden Balls” provides a neat opportunity to investigate the presence of reciprocal preferences outside of the laboratory and for substantial stakes. In particular, in the first two rounds contestants cast votes to determine who has to leave the show. Each vote carries a significant weight due to the small number of contestants, and voting against somebody can be viewed as unkind, particularly when the other votes were cast against other players. If people indeed have reciprocal preferences, we would expect that a

check, we have therefore also added the latter to Model 4. The effect of this additional control variable is insignificant (p = 0.573; untabulated), confirming our interpretation.

[17] See, for example, Kahneman, Knetsch and Thaler (1986), Blount (1995), Brandts and Solà (2001), Kagel and Wolfe (2001), McCabe, Rigdon and Smith (2003), Charness and Rabin (2002), Cox (2002, 2004), Falk, Fehr and Fischbacher (2003, 2008), Offerman (2002), Charness (2004), de Quervain *et al*. (2004). For theoretical accounts, see, for example, Rabin (1993), Levine (1998), Dufwenberg and Kirchsteiger (2004), Falk and Fischbacher (2006), Cox, Friedman and Gjerstad (2007), and Gul and Pesendorfer (2010).

contestant who makes it to the final in spite of her opponent's vote against her, has a lower propensity to cooperate.[18]

Although the voting is anonymous, it is often straightforward to deduce who has voted against whom. If a contestant in Round 1 (Round 2) receives three (two) votes it is obvious that all others have voted against her and that she herself has voted against the contestant who received one vote. For the other possible distributions of votes, we can usually deduce the individual votes from the banter preceding the vote, or, in the case of a tie, from the discussion following the vote. In the banter leading up to a vote, contestants generally make abundantly clear whom they intend to vote against (possibly out of an attempt to coordinate voting with other contestants). In the case of a tie, contestants openly discuss whom they want to leave the program; if it was not already clear from the banter whom they had originally voted against, this post-vote discussion generally makes it apparent. This procedure allows us to determine a contestant's vote 95 percent of the time. For various reasons it is much more difficult to determine clear instances of someone going out of their way to be nice to another player, so we limit our analysis to negative rather than positive reciprocity.

Based on the voting information, we create a dummy variable entitled "Vote received from opponent", taking the value of one if a contestant received a vote from her final opponent and zero otherwise. If we could not establish whether a contestant received a vote from her final opponent, she is assigned the value of zero as well (exclusion of these cases does not change our results). Since contestants who receive votes often do not make it to the next round, relatively few contestants qualify: as displayed in Table 2.2, 5 percent (28 subjects) of the final-round contestants received a vote from their opponent.

Model 2.6 in Table 2.4 includes the new dummy variable. In line with the idea that people have reciprocal preferences, the likelihood of a contestant to cooperate with her final

[18] Studying voting patterns in the internationally successful TV game show "The Weakest Link", both Levitt (2004) and Antonovics, Arcidiacono and Walsh (2005) find evidence that people reciprocate against people who voted against them in past rounds. However, because players in this show have an incentive to vote off players that are more likely to vote against them, they cannot rule out that this strategic concern drives the effect.

opponent plummets by approximately 21 percentage points if this opponent has voted against her earlier in the game ($p = 0.019$). There are, however, three alternative explanations for such a behavioral pattern that are unrelated to a genuine preference for reciprocity. Although we cannot rule out that these alternative explanations explain part of the effect, they do not appear particularly strong.

First, the causality may not run from receiving a vote to cooperativeness, but the other way around, players voting against contestants with a less cooperative disposition. This would imply that cooperation is also related to the number of votes received from other players, which appears not to be the case ($p = 0.231$; untabulated).

Second, a contestant may like to match her opponent's choice for reasons other than reciprocal concerns, and interpret the earlier vote against her as a signal that her opponent dislikes her and will not cooperate. However, her interpretation would generally not be legitimate: players do not cooperate less with someone they voted against ($p = 0.403$; untabulated). Moreover, the next section finds little support for such expectational conditional cooperation.

Last, a contestant's lower propensity to cooperate with someone who voted against her may be out of reputation concerns ("I am not to be messed with") instead of an intrinsic preference for reciprocity. However, when asked to explain their choice after the final decisions, contestants never use this costless opportunity to strengthen their message and point to their reciprocal nature.

2.6 Expectational Conditional Cooperation

There is considerable evidence that many people have a preference for conditional cooperation, defined as the desire to match the cooperation of others. In laboratory and field experiments, about half of the subjects are more willing to cooperate if others do so as well (*e.g.,* Fischbacher, Gächter and Fehr, 2001; Frey and Meier, 2004). Conditional cooperation can arise from reciprocal preferences, but also for other reasons. Social norms or a desire for conformity might account for it, and, especially in the laboratory, egalitarian motives can often explain conditional cooperative behavior because equality in payoffs generally only arises if players coordinate on their level of cooperation, as is also the case in our show.

Experimental studies typically investigate conditional cooperation in settings where subjects have the possibility to condition their behavior directly on the behavior of others. In everyday life, such clear-cut conditioning is usually not possible, especially in one-shot situations. Conditional cooperation then has to be based on expectations about the behavior of others, and the degree of coordination would depend on the predictive power of available information and on whether and how this information is interpreted. A natural question is whether conditional cooperation can be observed when the conditioning is only on an expectation of cooperation rather than on actual cooperation.

In “Golden Balls” it is not possible for a contestant to condition directly on her opponent’s behavior since the two are playing a simultaneous move game. However, we can investigate whether contestants condition their behavior on factors that form reliable predictors of their opponent’s behavior. That is, we can investigate the joint hypothesis that players make rational forecasts of their opponent’s behavior *and* then condition their behavior on those expectations.

The first step in such an analysis is determining the factors that a contestant could use to form an expectation about their opponent’s likelihood of cooperation. One such factor is whether an opponent made a promise to “split”. While the literature on conditional cooperation is rather recent, literature investigating the role of communication and, especially, promises in social dilemma situations already pointed towards tendencies of conditional cooperation. In a meta-analysis of Prisoner-Dilemma experiments, Sally (1995), for example, finds that cooperation occurs more often when the other player makes an explicit though non-binding promise that she will cooperate. The combination of a preference for conditional cooperation and a reluctance to lie (*e.g.,* Charness and Dufwenberg, 2005, 2006; Gneezy, 2005) can explain why promises have such an effect: people like to cooperate if others do, and a promise is a reliable signal of others’ behavior if they have a reluctance to lie.

Table 2.4: Binary Probit Regression Results [2/2]

The table displays results from the Probit regression analyses of contestants' decisions to "split" (1) or "steal" (0) the jackpot in the Prisoner's Dilemma at the end of the British TV game show "Golden Balls". The opponent variables measure the demographic characteristics of the contestant's opponent and are defined similar to the contestant's own demographic variables. Other definitions are as in previous tables.

	Model 2.6		Model 2.7		Model 2.8		Model 2.9		Model 2.10	
Age	0.002	(0.387)	0.003	(0.283)	0.002	(0.457)	0.002	(0.372)	0.002	(0.355)
Gender (male=1)	-0.249	(0.001)	-0.292	(0.000)	-0.258	(0.001)	-0.252	(0.001)	-0.247	(0.001)
Race (white=1)	0.149	(0.079)	0.162	(0.077)	0.146	(0.078)	0.147	(0.089)	0.148	(0.083)
City (large=1)	-0.034	(0.467)	-0.027	(0.552)	-0.035	(0.463)	-0.034	(0.463)	-0.037	(0.433)
London (London=1)	0.041	(0.565)	0.039	(0.540)	0.050	(0.501)	0.042	(0.551)	0.037	(0.597)
Education (high=1)	0.088	(0.068)	0.089	(0.041)	0.089	(0.067)	0.089	(0.065)	0.089	(0.066)
Student (student=1)	0.001	(0.988)	-0.025	(0.768)	0.008	(0.924)	-0.002	(0.983)	0.001	(0.988)
Age x Gender	0.011	(0.000)	0.008	(0.010)	0.010	(0.009)	0.011	(0.001)	0.011	(0.001)
Actual stakes (log)	-0.050	(0.000)	-0.054	(0.000)	-0.052	(0.000)	-0.051	(0.000)	-0.052	(0.000)
Potential stakes (log)	0.183	(0.004)	0.174	(0.004)	0.170	(0.006)	0.180	(0.004)	0.186	(0.004)
Transmissions	-0.000	(0.660)	-0.000	(0.106)	-0.000	(0.608)	-0.000	(0.584)	-0.000	(0.497)
Potential stakes x Transmissions	-0.001	(0.026)	-0.001	(0.022)	-0.001	(0.030)	-0.001	(0.025)	-0.001	(0.029)
Reciprocal Preferences										
Vote received from opp. (yes=1)	-0.215	(0.019)	-0.237	(0.015)	-0.202	(0.020)	-0.214	(0.026)	-0.216	(0.025)
Exp. Conditional Cooperation										
Promise (promise=1)			0.311	(0.000)						
Promise opp. (promise=1)			-0.080	(0.053)						
Age opp.					0.001	(0.732)				
Gender opp. (male=1)					-0.118	(0.140)				
Race opp. (white=1)					-0.026	(0.775)				
City opp. (large=1)					0.051	(0.288)				
London opp. (London=1)					0.027	(0.713)				
Education opp. (high=1)					0.116	(0.017)				
Student opp. (student=1)					0.069	(0.442)				
Age opp. x Gender opp.					0.004	(0.279)				
Past Deceitful Behavior										
Lie Round 1 opp. (lie=1)							-0.013	(0.782)		
Lie Round 2 opp. (lie=1)							-0.027	(0.578)		
Lie cash ball Round 1 opp. (lie=1)									0.037	(0.491)
Lie cash ball Round 2 opp. (lie=1)									-0.064	(0.295)
Lie killer ball Round 1 opp. (lie=1)									-0.055	(0.330)
Lie killer ball Round 2 opp. (lie=1)									0.007	(0.894)
Wald chi^2 (df)	59.65(13)		98.03(15)		62.40(21)		59.19(15)		62.72(17)	
Log pseudo-likelihood	-365.86		-343.31		-359.26		-365.64		-364.52	
McFadden R^2	0.078		0.135		0.095		0.079		0.082	
N	574		574		574		574		574	
Number of clusters	287		287		287		287		287	

In the conversation prior to the decision to either "split" or "steal", many contestants explicitly promise to "split" or otherwise make a definitive statement of their intention to do so. Based on the statements made in this small talk, we create a dummy variable labeled "Promise", indicating whether the contestant made an explicit, unambiguous promise or announcement that she will choose "split" (1) or not (0).[19] As shown in Table 2.2, about half (53 percent) of the contestants make such a promise. We investigate both whether observing a promise is predictive of the cooperative behavior of the contestant making the promise, and whether a contestant conditions her behavior on whether or not her opponent made a promise.

As shown in Table 2.4, Model 2.7, a player's promise is a highly significant predictor of her propensity to cooperate (p = 0.000). Those who make a promise are about 31 percentage points more likely to cooperate. In fact, an explicit promise is the single most reliable predictor of whether someone will cooperate.[20]

While a promise is a strong signal of cooperation, contestants whose opponent made a promise do not have a higher propensity to choose split. In fact, as Model 2.7 also shows, if an opponent promises to be cooperative, the other player even displays a marginally significant *decrease* in the likelihood of choosing "split". Belot, Bhaskar and van de Ven (2010a) obtain a similar result.

An explicit promise is the strongest predictor of cooperation, but, as we have previously shown, there are also demographic factors that a contestant could use to forecast cooperation. For example, we have seen that young males cooperate less than young females. However, inferences from this sort of analysis have to be tentative since there

[19] If a contestant responds affirmative to a question whether she will choose "split" or if she announces that she will not choose "steal", Promise takes the value of one as well. The value is zero in all other cases, including when people give the impression that they plan to split but do not explicitly express themselves as such, when they just refer to earlier intentions (for example, "I came here to split"), when they confine themselves to statements like "you can trust me" and "I will not let you down", and when they only express their preference for a coordinated outcome ("I want us to split"; "I do not want both of us to go home empty-handed").

[20] Of course, we do not interpret the promise as causing the cooperation. The direction of the causation could go the other way.

could be an additional confound if opponents have a taste for cooperating with someone with particular demographic characteristics.

As Table 2.4, Model 2.8 shows, we find little evidence that contestants condition their behavior on their opponents' background characteristics. The only weak evidence for conditional cooperation is that people cooperate significantly more frequently with higher educated opponents ($p = 0.017$), but education is not a very strong predictor of behavior. If we assess the joint significance of the various opponent background characteristics, we also find that they collectively do not have a significant effect on cooperation.[21] Analyzing the show "Friend or Foe?", Oberholzer-Gee, Waldfogel and White (2010) find no conditioning on opponent background characteristics in the first season, but they do find it in later seasons. They interpret this as conditional cooperation on the basis of learned expectations. We too have investigated whether conditioning arises as more episodes were transmitted, but we find no indication for such an effect.

As an alternative to the two models discussed above, we also examined a two-step approach. We first estimated each opponent's propensity to cooperate given her background characteristics and promise behavior, and then added the estimated propensity of opponents as an explanatory variable to our regression model. Again, we found no indication of conditionally cooperative behavior.

In summary, we find no evidence of expectational conditional cooperation. Apparently, either players cannot or do not forecast the behavior of their opponents, or they do not have conditionally cooperative preferences. Our evidence for reciprocal preferences in the previous section hints that it is the former rather than the latter interpretation that underlies this result.

Belot, Bhaskar and van de Ven (2010b) also provide evidence that predicting one's opponent's behavior is difficult. They had subjects watch clips from the Dutch

[21] We also looked at more complex mechanisms related to the similarity of the contestant's own background characteristics and those of her opponent, such as whether people cooperate more with those who are more similar to them ("social-distance" effects). In our data, there is no evidence of such behavior.

counterpart to "Golden Balls" and asked them to assess the likelihood of each contestant's cooperation. While the estimated likelihood for cooperators was significantly higher than for defectors, the difference was only seven percentage points.

2.7 Past Deceitful Behavior

In this section, we investigate whether lies influence opponents' willingness to cooperate. In the early rounds of the show, contestants have numerous opportunities to lie about the values on their hidden balls, lies that are quickly revealed to everyone. These lies can be consequential. If someone hides low value and killer balls and in so doing manages to remain in the game, she will have reduced the potential payoff to the remaining contestants.

In the final, contestants might be less likely to cooperate with opponents who have lied, either out of reciprocal concerns (*e.g.,* Brandts and Charness, 2003) or because they interpret lying as evidence of a self-interested nature and a sign of an imminent "steal" decision (*e.g.,* van Lange and Kuhlman, 1994). Thus, past deceitful behavior is not a separate possible determinant of cooperation, but rather a special case of either reciprocity or conditional cooperation, or both.

We collected data on the statements made by contestants and the actual values of the balls that they possessed, allowing us to specify various measures for deceitful behavior. The analyses reported here are restricted to the use of dummy variables. We have also tried more complex, continuous variables for lying, but these approaches yielded similar results.

We apply separate variables for each game round. The general variables take the value of one if the contestant lied, irrespective of whether she overstated the monetary value of a cash ball or failed to disclose a killer ball. To distinguish between these two types of lies, we also use specific variables for each type separately. It is not obvious which type of lie would be considered more objectionable. On the one hand, lying about killers is much more harmful to others than exaggerating the value of a cash ball, and, assuming a preference to reciprocate, doing so could then be expected to have a greater negative effect on an opponent's propensity to cooperate. On the other hand, lying about killer balls might also be more understandable, since killer balls have a much greater impact on a contestant's chances to be voted off the show. Players may realize that nearly everyone will fail to disclose a hidden killer ball, and thus not be inclined to punish such behavior. Lying about cash ball values might be more like gratuitous lying and be viewed more harshly, and, consequently, have a greater effect on cooperation.

As shown in Table 2.2, lying is rather common on the show: 41 percent of the contestants who made it to the final lied about their back row balls in Round 1, while 36 percent lied in Round 2 (some did both). Furthermore, in the first round, 24 percent overstated the value of a cash ball, while 21 percent hid a killer ball (some did both). For the second round, these figures are 15 and 24 percent, respectively.[22]

Table 2.4 displays the regression results when we add the general dummy variables (Model 2.9) and the dummy variables that distinguish between lying about killer balls and lying about cash balls (Model 2.10). We find that past lies of an opponent do not affect a contestant's propensity to cooperate: each of the six variables is insignificant ($0.295 < p < 0.894$). In addition to these simple tests, we also investigated whether lying is considered less fair and has more impact the more it is unexpected or "abnormal" given the circumstances, but again we found no significant effect.[23] Lying neither predicts a contestant's own cooperative behavior.[24] One plausible interpretation of these results is

[22] Conditional on having at least one killer ball on their back row, contestants hid a killer ball 50 (43) percent of the time in Round 1 (Round 2).

[23] We used a two-stage procedure to express the abnormality of a lie. For each round, we estimated a regression model that explains a contestant's propensity to lie, given the ball values on her back row, the ball values on her front row, and the rank of her front row balls relative to those of the other players. For each final contestant, the "abnormality" of a lie we then measured as the difference between unity and this estimated lie propensity.

[24] Such a relation might be expected if the propensity to be honest and the propensity to cooperate are influenced by a similar preference for "pro-social", "kind" or "fair" behavior. It has, for example, been argued that the reluctance to lie is driven by guilt aversion (Charness and Dufwenberg, 2005, 2006; Gneezy, 2005), and empirical analysis suggests that guilt aversion is also a strong driving force behind cooperative behavior (Dufwenberg, Gächter and Hennig-Schmidt, 2010).

that lying is seen as an inherent part of “Golden Balls” and therefore unobjectionable behavior, much as bluffing is considered in the game of poker (Charness and Dufwenberg, 2005).

2.8 Conclusions and Discussion

“Golden Balls” provides us with the possibility to examine cooperative behavior outside the conventional context of the laboratory with large sums of money at stake.

Our results provide support for the view that attitudes are strongly influenced by context. We find unusually high rates of cooperation when the luck of the game reduces the stakes to “merely” a few hundred Pounds. Such amounts are tiny in the light of the thousands and even tens of thousands the game is often played for, but would be considered very large in any laboratory setting. In the early days of the show, when the contestants have not had an opportunity to watch the show on TV and are still learning what kind of stakes are to be expected, cooperation rates appear to be influenced by the salient but normatively irrelevant value representing the maximum they could have been playing for with a lucky selection of balls. Over time, this effect vanishes, suggesting that expectations about stakes become well-informed.

We label the tendency to be unusually cooperative for what would normally be considered high stakes a “big peanuts” result. Players seem to feel that when making a choice about a few hundred dollars when they might otherwise have been dividing tens of thousands, they are playing for “peanuts”, and cooperate, perhaps thinking that it is not worth stealing for what they perceive to be so little money. This finding reinforces a similar result for risk taking behavior in another game show, “Deal or No Deal”. In that context, where the stakes were even higher, amounts of money in the tens of thousands of dollars became perceived as peanuts, since hundreds of thousands of dollars had been on the line. These are very big peanuts indeed.

Using the interaction that occurs among contestants prior to the final, we also examined the effects that past opposition and lying have on cooperation. Using the votes we find evidence to support the view that people have reciprocal preferences. Contestants are less likely to cooperate if their opponent has tried to vote them off the show at an earlier stage of the game. Lying on the other hand has no significant effect. We investigated several measures, but none was significantly related to cooperation. Lying is evidently not frowned upon in "Golden Balls", perhaps because it is expected. The different impact of opposition and lying might be related to their different nature in this game. Voting is a directed and aggressive act towards one specific contestant. Lying, on the other hand, is an undirected and defensive act.

With "Golden Balls" we are also able to investigate an interesting, expectational form of conditional cooperation. Specifically, since explicit promises to cooperate are strong predictors of actual cooperation, we can see whether players are more likely to cooperate with someone who has made such a promise. We find no evidence of such behavior. More generally, we find that players do not appear to condition their choice of whether to cooperate on factors that predict the cooperation likelihood of their opponent. Players may lack the ability or ignore the possibility to reliably interpret information about the expected behavior of others, or they may not have a preference for matching the other's choice. Given our finding that people reciprocate votes against them, the former explanation seems more likely. For situations beyond the context of our game these results suggest that conditional cooperation is not a very important phenomenon, at the least when direct conditioning is not possible and people would need to form expectations about the behavior of others.

We conclude with some comments on the generalizability of our results. There are three primary concerns. First, selection procedures may have affected the average cooperation rate in our study. Subjects self-select into auditions, are then selected by the producer, and during the game they themselves have the opportunity to vote off opponents they would rather not play the final with. For some demographic variables, selection may perhaps also have affected the correlation with cooperation. Unfortunately, we cannot substantiate our intuition that such effects are negligible, nor could we have prevented them if they would exist. Note, however, that selection procedures are inevitable in any lab experiment or field setting. Moreover, the subjects in our sample vary widely in terms of their demographic characteristics and as a group they seem to resemble a (middle-class)[25] cross-section of the general population more closely than subjects in most conventional experiments.

Second, subjects' behavior in a game show might be influenced by what could be called "a drive to win the contest". However, an important but hard to answer question would then be what "winning" actually means in this context. After all, it seems like a matter of personal social preferences whether winning is equivalent to a successful stealing attempt or to a successful coordination attempt.

Last, our contestants are not strictly playing a one-shot game. In the setting we study decisions are made on national TV, under the scrutiny of a studio audience and millions of viewers. This undoubtedly influences the behavior we observe. However, we do not feel that these special circumstances render our findings less interesting or less predictive of behavior in other settings. The truth is that every setting is, in some way, special. Subjects in a laboratory experiment know that their behavior is being scrutinized to some extent as well. Field settings are also "special"; bargaining over the price of a car or a house is different from negotiating compensation with a new employer or the division of household chores with a spouse. Although it would be fascinating for researchers to be able to surreptitiously study the outcomes of these sorts of interactions from the "real world", the researchers would still only be able to speculate on how their results would generalize to different real world settings. TV game shows offer a unique opportunity to study theoretically interesting behavior at stakes that are impossible to replicate in the lab. How the results compare with other contexts will be determined by future research.

In the absence of an ability to conduct such surreptitious field experiments in many domains, researchers are left with two alternatives: run experiments in the lab or the field, or study naturally occurring behavior in an interesting setting. This chapter is an

[25] For whatever reason, whether it is the interest in applying or the preferences of the producers, contestants are rarely very rich or very poor.

example of the latter strategy. Although a game show may seem like a strange environment, we think it may be closer to the situations that occur in the workplace than many other settings in which cooperation has been studied. Co-workers often must choose whether to cooperate, and their actions are often at least semi-public.

Finally, the big peanuts phenomenon, perhaps the most interesting finding in this chapter, is one that does not appear to depend in any important way on the specific game show environment. As a US Senator once famously said, “a billion here, a billion there, pretty soon you’re talking real money.”

Chapter 3 | Standing United or Falling Divided? High Stakes Bargaining in a TV Game Show

In this chapter, we examine high stakes three-person bargaining in a game show where contestants bargain over a jackpot that is split into three unequal shares and ranges from about $10,000 to $185,000. In contrast to the commonly held view that fairness concerns will be unimportant when monetary incentives are sufficiently high, we find that individual behavior and outcomes are strongly influenced by equity concerns: those who contributed more to the jackpot claim larger shares, are less likely to make concessions, and take home larger amounts. Threatening to play hardball is ineffective. Although contestants who announce that they will not back down do well relative to others, they do not secure larger absolute amounts and harm others. In addition, there is no evidence of a first-mover advantage and little evidence that demographic characteristics matter.

This chapter is based on the paper "Standing United or Falling Divided? High Stakes Bargaining in a TV Game Show", co-authored by Martijn J. van den Assem, Colin F. Camerer, and Richard H. Thaler (van Dolder *et al.*, 2013). We thank developer and format holder Talpa for granting the right to use copies of *Divided* for our study, and producer Endemol UK and in particular Tara Ali, Gillian Bristow and Marie-Josee Grenier for providing us with recordings and background. The chapter benefited from discussions with seminar participants at the University of Chicago Booth School of Business, the Erasmus University of Rotterdam and Maastricht University, and with participants of the Behavioral Decision Research in Management (BDRM) 2012 conference in Boulder, the Society for the Advancement of Behavioral Economics (SABE) 2012 conference in Granada, the Tilburg Institute for Behavioral Economics Research (TIBER) 2012 conference in Tilburg, and the Subjective Probability, Utility and Decision Making (SPUDM) 2013 conference in Barcelona.

3.1 Introduction

Bargaining is ubiquitous in our professional and private lives. In politics, negotiations form the basis of coalitions between political parties in multi-party systems and between countries in international relations. In business, negotiations determine wage compensation schemes for employees, the division of surpluses between trading firms, and the terms of mergers and acquisitions. In our private lives, negotiations underlie the division of household chores between our partner and ourselves, and the division of property after a divorce if we fail to do so properly. Not surprisingly, bargaining has received considerable attention in both economics and psychology.

Economists have been particularly interested in the efficiency and distribution of bargaining outcomes (Muthoo, 1999). In theoretical accounts, the emphasis is on models invoking stylized representations of bargaining settings that facilitate the derivation of equilibrium predictions, with Rubinstein's alternating-offers bargaining model being especially influential (Rubinstein, 1982). As a result of this emphasis, bargaining experiments in economics typically employ fixed bargaining protocols and are conducted anonymously using computer terminals. Psychologists have studied a much wider range of topics. Examples include people's perceptions about the bargaining situation and the possible outcomes, their behavior during the bargaining process, and the roles of values, culture and communication (Thompson, 1990; Bazerman *et al.,* 2000). In contrast to the stylized bargaining settings in economics, psychological experiments often have subjects participate in free-form face-to-face negotiations.

While bargaining has been extensively studied from various perspectives, most empirical evidence on bargaining behavior and outcomes derives from laboratory experiments. Real-world data generally entail a lack of control, making it difficult—if not impossible—to distinguish between competing hypotheses. It is still an open question, however, to what extent findings from the laboratory generalize to real-world environments (Levitt and List, 2007a, 2008; Camerer, 2011). One of the concerns arises from the fact that volunteering students are a non-random sample of the population. Also, experimenters are mostly unable to employ high stakes, begging the question whether or not results will generalize to situations of significant economic importance. In the present study, we examine high stakes bargaining using data from the British TV game show *Divided*. This setting has the unique property that it allows for the study of bargaining behavior in a controlled setting where the stakes are high, for a diverse subject pool.

There is a growing literature that uses game shows to study decision-making. TV shows can offer unique research opportunities, because contestants often face relatively well-defined choice problems for high stakes. Prior studies have focused on risky choice, strategic decision making, discrimination, and cooperative behavior (Bennett and Hickman, 1993; Gertner, 1993; Metrick, 1995; Berk, Hughson and Vandezande, 1996; Beetsma and Schotman, 2001; Tenorio and Cason, 2002; Fullenkamp, Tenorio and Battalio, 2003; Levitt, 2004; List, 2004a; Antonovics, Arcidiacono and Walsh, 2005; List, 2006; Post *et al.,* 2008; Belot, Bhaskar and van de Ven, 2010a; Oberholzer-Gee, Waldfogel and White, 2010; van den Assem, van Dolder and Thaler, 2012). The present chapter is the first to exploit the favorable combination of features of a game show to study bargaining.

In *Divided*, three contestants collectively build up a jackpot through answering general quiz questions. Across episodes, their jackpot ultimately ranges from approximately \$10,000 to \$185,000, and averages over \$50,000. In the second phase of the game, the team's accumulated money amount is divided into three unequal parts of, for example, 60, 30 and 10 percent. Contestants in turn have to claim one of these shares. If they do not immediately agree on who takes which share, they have 100 seconds to negotiate and reach consensus. With each second they take they lose one percentage point of the initial jackpot, and after 100 seconds there is nothing left. This final stage can thus be seen as a natural bargaining experiment where the "subjects" have to unanimously decide on the allocation of three indivisible shares, in a format that allows face-to-face communication and incorporates (close to) continuous costs to bargaining.

Overall, we find that 10 percent of the teams agree immediately, 71 percent do so while the timer counts down, and 19 percent fail to reach agreement and go home empty-handed. The efficiency rate, or the average fraction of the jackpot that is actually awarded, is approximately 50 percent.

Because the jackpot is determined by teams' answers to trivia questions, we are able to investigate the influence of entitlements on bargaining behavior and outcomes. In real world settings, entitlements can potentially arise from a wide range of sources, including history, custom, status quo, and contributions (Gächter and Riedl, 2005). In *Divided*, however, the only and apparent source of entitlements are contestants' individual contributions to the communal jackpot. Equity theory suggests that contestants prefer outcomes to be proportional to inputs (Homans, 1958; Adams, 1965; Walster, Berscheid and Walster, 1973). To the best of our knowledge, all tests of equity theory thus far have relied on survey data or experiments employing relatively low stakes. With our data, we are able to test equity theory in a controlled bargaining setting where the stakes are high.

We find that equity concerns play an important role in the bargaining process. Contestants who contributed more to the communal jackpot claim a larger share, are less likely to lower their claim during the bargaining process, and end up with a larger fraction of the jackpot.

At the start of the bargaining process, contestants have 15 seconds to make their case and stake their claim to one of the shares. About one in four use this opportunity to make a hardball announcement, by adding a statement to their initial claim that they will not back down from it. We find that contestants who use this threat do well relative to others. However, as a result of increased bargaining costs, hardball announcements do not generate higher earnings in an absolute sense and lower the earnings of others.

The effect of the stakes on behavior in our data is twofold. First, when the stakes are relatively high, contestants are more likely to make a hardball announcement. Second, the effect of the stakes on the likelihood of concessions is U-shaped: concessions occur relatively often at the low and high stake levels, and less so in between. These two effects together suggest that at some point, for higher stakes, the increased bargaining costs are outweighing the increased incentive to try and fight for the top share and the possible disutility of being worse off than opponents.

Finally, we obtain a number of potentially interesting null results. There is no evidence of a first-mover advantage: the order in which contestants are to make their initial claims has no influence on these claims and is neither related to contestants' subsequent behavior nor to their payoffs. Furthermore, we find little evidence that behavior or outcomes are related to demographic characteristics including age or gender.

The chapter proceeds as follows. In Section 3.2, we describe the game show in greater detail and discuss our data material. Section 3.3 explains the various explanatory variables that we use and discusses related literature. Section 3.4 presents our analyses and results. Section 3.5 discusses the results and concludes.

3.2 Game Show and Data

Description of Divided

The format of *Divided* was developed by the Dutch media firm Talpa, and produced for the ITV network in the United Kingdom by Endemol UK. The show debuted on British TV in May 2009 and ran until May 2010. A total of 53 episodes were aired, and, at the time of writing, no further episodes were aired thereafter.

Each game is played with three contestants who are strangers to each other, and consists of two stages: one in which the contestants team up to accumulate a communal jackpot through answering quiz questions, and one in which they have to divide the jackpot between them.

The first stage lasts for a maximum of five rounds. Round 1 has five questions that are worth up to £3,000 each.[26] In the subsequent four rounds the number of questions and the maximum value per question are 4, 3, 2, 1 and £7,500, £15,000, £30,000, £75,000,

[26] Values in British pounds can be translated into US dollars using a rate of $1,60 per pound, an approximate average of the exchange rate during the period in which the show ran.

respectively.[27] In theory, the maximum potential jackpot is £225,000. How much a question actually contributes to the jackpot depends on the team's speed of answering: they have one hundred seconds to agree on an answer, and with each second that passes the value of a correct answer falls by one percentage point of its initial value. Incorrect answers halve the jackpot, and after three mistakes the team is out of the game. At the end of each round, the team can decide to stop and divide the jackpot, but only if they make that decision unanimously and within 15 seconds – otherwise, the next round starts automatically. Figure 3.1 presents a schematic overview of the first stage of the game.

The second stage comprises the bargaining element of the game that is central to our analysis. The jackpot is split into three unequal shares representing, for example, 60, 30 and 10 percent of the total jackpot. The largest prize is marked share A, the middle B, and the smallest C. The players unanimously have to decide who gets which. First, they each receive 15 seconds to make their case and stake their claim to one of the shares. The order in which they are asked to do so is determined by their positions on the stage (from left to right, for viewers). If they do not agree immediately, they have 100 seconds to reach consensus in a free-form discussion. With each second that passes before they agree they lose one percentage point of the initial jackpot, and after 100 seconds there is nothing left. Halfway, after 50 seconds, there is a time-out. In this brief pause, the contestants keep silent and the game show host summarizes the situation by bringing to their attention how much they have lost and what is left, or by enumerating the remaining values of the three different shares.

[27] Most questions are general knowledge questions of the multiple choice type with one out of three answers being correct (such as "which of these flags is the flag of the Netherlands"). In some cases, three alternatives need to be put in a particular order (e.g., "starting with the youngest, put these actresses into age order"). In the fifth round, the question can have multiple correct alternatives and out of a list of three the team must select all of them (e.g., "which of these countries has a currency named the pound").

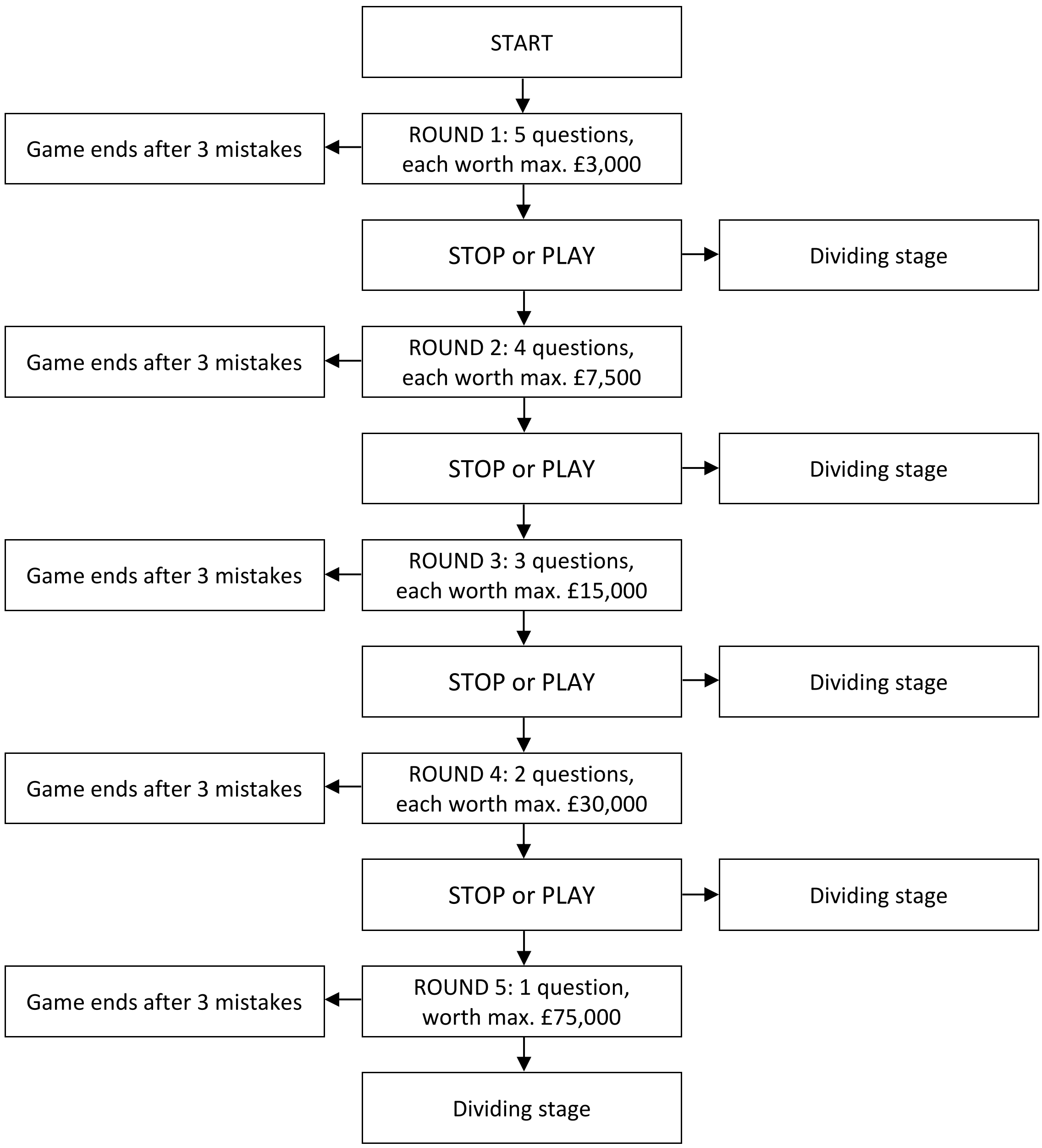


Figure 3.1: Flow Chart of the First Stage of the Game

Three contestants first play a maximum of five rounds of quiz questions in which they team up to accumulate a jackpot. Correct answers increase the jackpot, while incorrect answers halve it. A third mistake ends the game, and all contestants then leave empty-handed. At the end of each of the first four rounds, the team can voluntarily decide to proceed to the second stage. In this final part of the game they have to divide the money they accumulated between them.

The producer has applied a standard procedure to select contestants. A spokeswoman of Endemol informed us that anyone could apply to be on *Divided* by submitting a detailed application form. Shortlisted contestants were then invited to an audition in order to determine their skill in playing the game, their character, and their suitability to appear on a TV game show. Producers watched tapings of these auditions and put together teams of three that they deemed to be "good mixes of characters". Contestants and teams were thus not randomly drawn from the general population, but at the same time the selection process does not seem to create any obvious confounds in our analyses.

Data and Summary Statistics

With permission from Talpa we received copies of all episodes from producer Endemol UK. For each episode we collected data on the relevant observables in the show, including demographic characteristics of the contestants, the results for each quiz question and the individual contributions to the team's answers, the decisions at the end of each round to play on or stop and divide the money, contestants' claims and how these changed during the bargaining phase, whether and when agreement was reached, and the individual payoffs.

Combined, the 53 episodes comprise the games of 56 teams, with some starting in one episode and continuing their game in the next. Because 13 teams leave the show early after three incorrect answers, only 43 games are used in our analyses. In terms of observable demographic characteristics, the composition of the eliminated teams is not significantly different from those who did play the bargaining game. Most of the 43 included teams are mixed-gender teams. Only in 4 exceptions the three contestants are all female. Men and women each represent half of the contestant pool. The average contestant is 36 years of age, with the youngest being 18 and the oldest 70. The large majority (94%) are white, a small minority (9%) are students, about half (49%) are from a conurbation with a population exceeding 250,000 inhabitants, and the majority (91%) are from outside the area where the recording studios were located (Greater Manchester).

Table 3.1 displays some descriptive characteristics for *Divided*. Quiz questions become harder with each round. Even though more skillful teams are more likely to play the later rounds, the fraction of correct answers monotonously decreases from 90 percent in Round 1 to only 29 percent in Round 5. Out of the thirteen teams not reaching the bargaining stage, seven are eliminated in Round 3 (and one or two in each of the other four rounds). On average, successful teams decide to start the bargaining over the division of the jackpot after three rounds of trivia questions. Mostly, playing on would have implied a high risk of losing it all: the modal number of incorrect answers when teams voluntarily decide to move to the second stage is two (72%).

On average, the final jackpot is £33,512. The variation is large: from £7,282 to £115,755. In two-thirds of all cases the three shares in the jackpot – A, B, and C – represent close to 60, 30, and 10 percent, respectively. Only two other subdivisions occur: 70/20/10 and 65/25/10, both in 16 percent of the cases. The smallest share thus always represents about ten percent of the prize money, but the largest and the middle share vary in size. The average initial value per jackpot share of £11,171 is many times the amounts typically used in laboratory experiments, and also a large sum relative to the median gross weekly earnings of £404 in the UK in April 2010 (Office for National Statistics 2010). Overall, 58 percent of the shares exceed three months of median UK earnings, and 13 percent are even larger than the median annual salary.

Most contestants initially claim the largest share: 79 percent opt for A, 16 percent pick B and only 5 percent content themselves with C straight away. As a result, the three contestants all claim A about half the time (51%). Only four teams (9%) agree immediately. In the end, 22 teams (51%) manage to reach agreement within 50 seconds and 35 (81%) reach agreement before the timer has counted down and all the money is gone. Eight (19%) fail completely and go home empty-handed. Including the teams that agree immediately or leave empty-handed, the average bargaining process lasts 50 seconds. Correspondingly, the average efficiency rate amounts to 50 percent. Figure 3.2 displays the distribution of the bargaining duration.

Table 3.1: Selected Game Show Characteristics

The table shows selected characteristics for the British TV game show *Divided*, extracted from our sample of 53 episodes. *Answer in Round r* (*r* = 1, 2, … ,5) is the status of the team's answer to a question in Round *r*, with a value of 1 (0) for a correct (incorrect) answer. *Jackpot change Round r* (*r* = 1, 2, … ,5) records the difference between the size of jackpot at the end and at the start of round *r* for all teams still in play at the end of the round. *Quiz rounds* measures the number of quiz rounds completed before elimination or entering the bargaining stage. *Mistakes* is the accumulated number of incorrect answers when the team enters the bargaining stage. *Jackpot* describes the size of the jackpot. *Prize A* (*Prize B*, *Prize C*) */ jackpot* expresses the size of the largest (middle, smallest) share as a fraction of the jackpot. *Initial claim* indicates the share that the contestant claims before the timer starts counting down, with a value of 3 (2, 1) for A (B, C). *Final claim* is the share that the contestant claims at the end of the bargaining process, with a value of 3 (2, 1) for A (B, C). *Resolution before t=0* (*t=50*, *t=100*) is a dummy variable taking the value of 1 if the team reaches agreement before the timer starts (before 50 seconds have passed, before 100 seconds have passed). *Time to resolution* measures the duration of the bargaining process in seconds. *Prize won* (*if non-zero*) records the prize the contestant takes home (if she did not leave empty-handed). *Prize won* (*if non-zero*) */ initial jackpot* records her prize as a fraction of the initial jackpot (if she did not leave empty-handed). All monetary values are in UK Pounds (£1.00 ≈ $1.60).

	N	Mean	Stdev	Min	Median	Max
All teams						
Answer Round 1 (correct=1)	280	0.90	0.30	0.00	1.00	1.00
Answer Round 2	219	0.87	0.33	0.00	1.00	1.00
Answer Round 3	119	0.68	0.47	0.00	1.00	1.00
Answer Round 4	37	0.65	0.48	0.00	1.00	1.00
Answer Round 5	7	0.29	0.49	0.00	0.00	1.00
Jackpot change Round 1	55	9,010	3,135	2,963	9,360	13,170
Jackpot change Round 2	54	14,170	8,112	-5,648	16,125	25,500
Jackpot change Round 3	34	9,665	19,762	-25,342	5,293	37,950
Jackpot change Round 4	17	5,698	31,528	-51,919	-2,280	53,400
Jackpot change Round 5	5	-6,319	28,695	-41,040	-17,887	27,750
Teams eliminated after three mistakes						
Quiz rounds	13	2.23	1.09	0.00	2.00	4.00
Teams playing bargaining stage						
Quiz rounds	43	3.16	1.00	2.00	3.00	5.00
Mistakes	43	1.70	0.51	0.00	2.00	2.00
Jackpot	43	33,512	26,154	7,282	23,288	115,755
Prize A / jackpot	43	0.62	0.04	0.59	0.60	0.70
Prize B / jackpot	43	0.27	0.04	0.19	0.30	0.30
Prize C / jackpot	43	0.10	0.00	0.10	0.10	0.12
Initial claim (A=3, B=2, C=1)	129	2.74	0.53	1.00	3.00	3.00
Final claim (A=3, B=2, C=1)	129	2.14	0.83	1.00	2.00	3.00
Resolution before t=0 (resolution=1)	43	0.09	0.29	0.00	0.00	1.00
Resolution before t=50	43	0.51	0.51	0.00	1.00	1.00
Resolution before t=100	43	0.81	0.39	0.00	1.00	1.00
Time to resolution (in seconds)	43	50.26	35.39	0.00	50.00	100.00
Prize won	129	5,633	8,616	0	2,615	56,895
Prize won if non-zero	105	6,921	9,075	135	4,030	56,895
Prize won / initial jackpot	129	0.17	0.18	0.00	0.10	0.66
Prize won if non-zero / initial jackpot	105	0.20	0.17	0.01	0.15	0.66

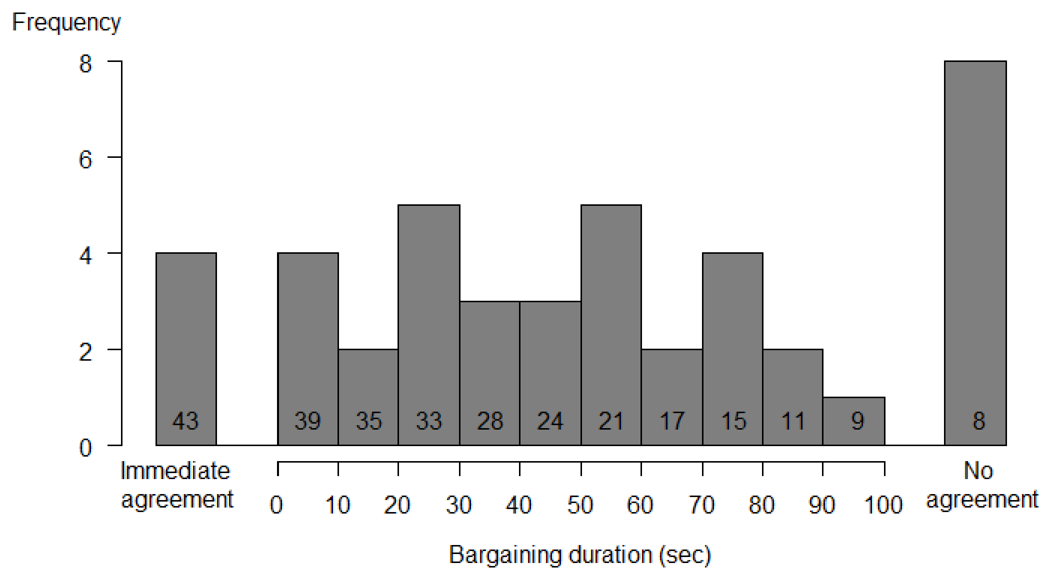


Figure 3.2: Bargaining Duration

The histogram shows the distribution of bargaining duration for the 43 teams in our sample, where the time frame is divided into ten-second intervals. The leftmost (rightmost) bar corresponds to the teams that reach immediate agreement (fail to reach agreement). The number of teams not yet in agreement immediately prior to a given duration category is displayed at the bottom of the bar.

The stakes-weighted efficiency rate obtained by dividing the sum of earnings across all games by the sum of all jackpots is also equal to 50 percent. On average, a contestant who plays the bargaining game goes home with £5,633. The 105 contestants who end up with a non-zero prize take home £6,921 on average, with a median of £4,030. Would we have run this show as an experiment ourselves, the total costs in subject payoffs alone would have been £726,706, or about $1.16 million.

3.3 Variables of Interest and Background

Demographic Characteristics

Psychologists have devoted considerable attention to studying individual differences in negotiation, especially during the 1970s and the early 1980s. The general picture arising from studies into the roles of demographic and personality characteristics is one of contradictory findings, frequent null results, and low explanatory power (Rubin and Brown, 1975; Thompson, 1990). For gender, meta-analyses indicate that males are more competitive in bargaining (Walters, Stuhlmacher and Meyer, 1998) and better in acquiring favorable outcomes (Stuhlmacher and Walters, 1999), but the differences are slim and sensitive to the specific experimental conditions employed.

A more recent study by Elfenbein *et al.* (2008) does show substantial individual differences in bargaining performance between individuals. They had subjects participate in multiple negotiations with different counterparts, and find that individual differences are persistent but unrelated to a wide range of personality and background variables. Bowles, Babcock and McGinn (2005) study the role of structural ambiguity, defined as the degree of uncertainty regarding the economic structure of the negotiation. They find that gender differences are only present under a high level of structural ambiguity. Such ambiguity is largely absent in our show. Altogether, prior work suggests that it is not very likely that there are strong effects of demographic characteristics in our data.

The demographic variables that we study are gender, age and education. *Gender* is a dummy variable that takes the value of one if the contestant is male, *Age* is a continuous variable measuring the contestant's age in years, and *Education* is a dummy variable that takes the value of one if the contestant has at least a bachelor degree. Contestants normally mention their age when they introduce themselves at the start. By contrast, they generally do not talk about their education during the show. We therefore estimate their level of education on the basis of their occupation and other information they provide. Contestants who are currently enrolled in higher education and contestants whose job title suggests work experience equivalent to the bachelor level or higher are also included in the higher education category. The proper binary values are generally clear.[28] We have also distinguished between student/non-student, urbanite/villager and

[28] In eight exceptions, we had to estimate a contestant's age on the basis of her physical appearance and other information given in the introductory talk. Seven contestants provided no job or other relevant information that we could use to assess their education level; we included these in the lower education category.

white/non-white contestants, but omit these characteristics from our analyses. There are relatively few students (9%) and non-white contestants (6%), and the results for the three variables would be insignificant throughout.

Entitlement Measures

Entitlements are subjectively-held fairness judgments that people perceive as rights they wish to defend, and can arise from history, custom, the status quo, or from the contributions that people made to the bundle that has to be divided (Schlicht, 1998). In an experiment, Gächter and Riedl (2005) show that such entitlements influence bargaining behavior and outcomes in the absence of legal property rights.

In our show, the three contestants similarly have no legal rights to any of the shares. Still, they may feel entitled to a certain share of the jackpot due to their contributions. Sociologists and social psychologists have stressed the role of equity as a criterion for distributive justice in situations of social interaction (Homans, 1958; Adams, 1965; Walster, Berscheid and Walster, 1973). Equity theory states that outcomes are only fair if they are proportional to inputs. Imagine two actors, *A* and *B*, and denote their outcomes by *O* and their inputs by *I*, then according to the equity formula a distribution is fair if $O_A/I_A = O_B/I_B$. Empirical studies have largely confirmed the idea that people care about equity. Many show that inputs and outcomes are positively related (Konow, 2003), and some even demonstrate that fairness judgments follow the exact proportionality of inputs and outputs posited by equity theory (Schokkaert and Overlaet, 1989; Konow, 1996; Clark, 1998; Konow, 2000).

To the best of our knowledge, our study is the first to investigate the role of equity concerns in a controlled environment with high stakes. Prior work used surveys or experiments with no or relatively low performance-based financial incentives. The high stakes of *Divided* are especially interesting in the light of the argument that fairness considerations will be unimportant if the stakes are sufficiently high (Rabin, 1993; Telser, 1995; Levitt and List, 2007a).

Unfortunately, there is not just one single way in which contestants' contributions to the jackpot can be objectively quantified in our game. In fact, there are numerous possibilities. As a result, different contestants may adopt different definitions, which could be detrimental to the explanatory power of individual measures. Moreover, the possible lack of consensus about contributions is potentially aggravated by self-serving bias in contestants' attributions (Camerer and Loewenstein, 1993; Loewenstein *et al.,* 1993; Babcock *et al.,* 1995; Babcock, Wang and Loewenstein, 1996; Babcock and Loewenstein, 1997). Consequently, even if contributions determine entitlements, we may not necessarily find strong correlations between contribution measures and behavior and outcomes.

The results we present are for relatively simple contribution measures, in which we credit (in)correct answers by the team to the players who argued for (against) the correct answer. We distinguish between a composite measure that combines the credits for correct and incorrect answers into one metric, and measures that isolate the contributions to correct and incorrect answers. We have also investigated various measures that account for the money won or lost with a specific question, but the results are insensitive to such alternative approaches.

More specifically, if the group gave a correct answer, the credit for this answer is divided equally over all contestants who argued in favor of it. For example, if all three contestants argued for the correct answer, then each contestant receives one-third of the credit. If two did so, then both receive half of the credit, and if only one argued for the correct answer she receives the full credit. Those who did not argue for any particular answer, argued for a wrong one, or argued for multiple answers (including or not including the correct one) receive no share of the credit.[29] If the group gave an incorrect answer, the credit is divided equally over those who argued in favor of one of the incorrect answers.

[29] There are three exceptions to this rule: (i) if all contestants argued both for and against the correct answer but managed to come to the correct answer together, they are each assigned one-third of the credit; (ii) if two contestants argued both for and against the correct answer and came to the correct answer together while the third remained silent, then these two share the credit; (iii) if contestants made a random guess and this guess turned out to be correct, then they share the credit.

Those who did not argue for any particular answer or argued for the correct one only are not assigned any credit in this case.

Based on these credits, we create the following contribution measures:

- *Contribution overall*: a contestant's overall contribution, calculated by adding up all credits of a contestant for questions answered correctly and subtracting all credits of a contestant for questions answered incorrectly. We normalize by dividing by the total number of correct answers minus the total number of incorrect answers of the team.

- *Contribution correct*: a contestant's contribution to the team's correct answers, calculated by adding up all credits of a contestant for questions answered correctly. We normalize by dividing by the total number of correct answers.

- *Contribution incorrect*: a contestant's contribution to the team's incorrect answers, calculated by adding up all credits of a contestant for questions answered incorrectly. We normalize by dividing by the total number of incorrect answers.

Situational Variables

In addition to the demographic and contribution variables discussed above, we also consider the influence of the stakes, the variance across the percentage shares to be divided, and the order in which contestants make their initial claims.

A priori, the effect of stakes can go either way. On the one hand, higher stakes give a stronger incentive to fight for the top share and so might lead to bigger claims and more impasses. On the other hand, the costs of bargaining are in direct proportion to the stakes, which might create an additional incentive to strive for a speedy resolution and to easily succumb to others' pressure if the stakes are high. In addition, contestants may experience (dis)utility from being better (worse) off than others and feeling victorious (deprived), and it is unclear how these non-monetary costs and benefits are traded off against the monetary costs and benefits of bargaining. For flexibility and ease of interpretation we use dummy variables representing the different quartiles of the stake distribution. (A quadratic specification yields similar results.)

Generally, consensus will be more difficult to achieve if the difference between the shares is larger. If share A increases relative to the other shares, contestants will have a stronger incentive to attempt to take this top share home. Furthermore, the distribution of the percentage shares tends to be more unequal than the distribution of the contributions. If contestants care about equity, and especially if they care about receiving at least what they deserve, larger differences between the shares are likely to make bargaining more difficult. As a measure for the divergence between the prizes we use the variance across the percentage shares (the standard deviation leads to similar results).

The order in which contestants express and motivate their initial claim can affect the bargaining process. By claiming share A, the first contestant may “force” the others to pick a smaller share if they wish to avoid an impasse and the risks and costs it entails. Alternating-offer bargaining models and experiments in economics point at the existence of such a strategic advantage for the first mover (Rubinstein, 1982; Sutton, 1986; Ochs and Roth, 1989). Although structures with offers and counteroffers are often considered intuitively appealing because they resemble most real-life negotiations (Muthoo, 1999), the strict alternating-offer protocol is a considerable abstraction. It is therefore interesting to investigate whether a first-mover advantage also occurs in our setting, where contestants’ initial claims are followed by free-form bargaining. We use a dummy variable that takes the value of one if the contestant was the first to make her case and stake her claim to one of the shares.

Claim Variables

Finally, we look at contestants’ announcements to play hardball by stating not to back down from their initial claim. We investigate the possible determinants of handball announcements, and examine the relations between such statements and actual behavior and outcomes. In his influential paper on bargaining, Schelling (1956) stresses the importance of commitment strategies. Later work has incorporated the notion of commitment into formal bargaining models (Crawford, 1982; Kambe, 1999; Muthoo, 1996; Abreu and Gul, 2000; Compte and Jehiel, 2002; Ellingsen and Miettinen, 2008). In our bargaining setting, contestants cannot formally commit themselves in the sense that they are always free to adjust their claim without incurring monetary costs. However, contestants may attempt to convince others that they feel internally committed to a specific share. In our analyses of the effect of hardball announcements we control for contestants' initials claims.

3.4 Analyses and Results

In this section, we first analyze the determinants of contestants' initial claims. Next, we examine the correlates of hardball announcements and actual concession behavior. Last, we analyze the factors driving bargaining outcomes. Table 3.2 presents descriptive statistics for variables not yet included in Table 3.1.

Initial Claims

Table 3.3 shows the ordered probit regression results for contestants' decisions to initially claim share A (3), B (2) or C (1). We find that a contestant's contribution to the jackpot determines the share that she chooses: those who contributed more are significantly more likely to claim a large share. Interestingly, the results for Model 3.2 suggest that there is an asymmetry between the types of contributions. When we distinguish contributions to correct answers from contributions to incorrect answers, we find that the effect is driven by the positive contributions only. The influence of a contestant's role in mistakes is insignificant. Neither the demographic characteristics nor the situational variables influence the initial claims.

Due to the inherent nonlinearities of a probit model, the sizes of the coefficients are not easy to interpret. Transformed into marginal effects at the covariate means, we find that a 10 percentage-point increase of *Contributions overall (correct)* increases the likelihood of claiming share A by 8.2 (11.1) percentage points, and decreases the likelihood of claiming share B or C by 6.0 (8.2) and 2.2 (2.9) percentage points, respectively.

Table 3.2: Descriptive Statistics

The table shows descriptive statistics for our sample of 129 contestants who bargain over their share of the jackpot in the final stage of the British TV game show *Divided*. *Age* is the contestant's age measured in years. *Gender* is a dummy variable taking the value of 1 if the contestant is male. *Education* is a dummy variable taking the value of 1 if the contestant has completed or is enrolled in higher education (bachelor degree or higher) or has equivalent working experience. *Variance shares* denotes the variance across the three percentage shares to be divided. The contribution variables measure the contestant's entitlement to the communal jackpot. *Contribution overall* measures her contribution across all quiz questions. *Contribution correct* (*incorrect*) measures her contribution to the team's correctly (incorrectly) answered questions only. *Announce hardball, Opp. announce hardball* and *Concession* are dummy variables taking the value of 1 if the contestant stated not to back down from her initial claim, faced at least one opponent who had stated not to back down, or gave in during the bargaining process, respectively. *Concession* is not defined if the team agrees immediately or if the contestant initially picked share C. All monetary values are in UK Pounds (£1.00 ≈ $1.60).

	N	Mean	Stdev	Min	Median	Max
Age	129	36.16	12.23	18.00	34.00	70.00
Gender (male=1)	129	0.50	0.50	0.00	0.00	1.00
Education (high=1)	129	0.30	0.46	0.00	0.00	1.00
Variance shares	129	0.05	0.01	0.04	0.04	0.07
Contribution overall	129	0.33	0.12	0.07	0.33	0.70
Contribution correct	129	0.33	0.09	0.10	0.33	0.56
Contribution incorrect	129	0.33	0.20	0.00	0.33	1.00
Announce hardball (hardball=1)	129	0.23	0.42	0.00	0.00	1.00
Opp. announce hardball (hardball=1)	129	0.30	0.46	0.00	0.00	1.00
Concession (concession=1)	115	0.50	0.50	0.00	0.00	1.00

Hardball Announcements and Concessions

When agreement is not reached immediately, contestants have to negotiate to determine who gets which share. During this negotiation, some will have to make concessions to bring agreement within reach. Here we will analyze the behavior of contestants during this negotiation process. First, we investigate what determines whether contestants make hardball announcements by stating not to back down from their initial claim. Second, we investigate the likelihood that a contestant actually makes a concession during the bargaining process.

Table 3.3: Ordered Probit Regression Results on Initial Claims

The table displays results from the ordered probit regression analyses of contestants' decisions to initially claim share A (3), B (2), or C (1) in the bargaining stage of the British TV game show *Divided*. *First mover* is a dummy variable taking the value of 1 if the contestant was the first to make her claim. The stakes quartile dummies are used as a flexible specification for the effect of stakes. Definitions of other variables are as inTable 3.2. Standard errors are corrected for clustering at the team level, *p*-values are in parentheses.

	Model 3.1		Model 3.2	
Demographic characteristics				
Age	-0.012	(0.233)	-0.013	(0.227)
Gender (male=1)	-0.187	(0.517)	-0.230	(0.428)
Education (high=1)	0.005	(0.988)	-0.008	(0.980)
Situational variables				
First mover (first=1)	0.018	(0.949)	0.007	(0.981)
Stakes 2nd quartile	0.279	(0.529)	0.288	(0.510)
Stakes 3rd quartile	0.082	(0.820)	0.090	(0.799)
Stakes 4th quartile	-0.235	(0.515)	-0.208	(0.558)
Variance shares	12.265	(0.313)	11.742	(0.335)
Contribution variables				
Contribution overall	3.007	(0.002)		
Contribution correct			4.133	(0.004)
Contribution incorrect			-0.660	(0.206)
α_1	-0.311	(0.654)	-0.218	(0.803)
α_2	0.649	(0.345)	0.750	(0.386)
Log pseudo-likelihood	-74.45		-73.71	
McFadden R^2	0.075		0.084	
N	129		129	
Number of teams	43		43	

The columns labeled Model 3.1 and Model 3.2 in Table 3.4 contain the results from the probit regression analyses of the decision to make a hardball announcement (1) or not (0) at the start of the bargaining stage. The two models are estimated for the subset of contestants who initially claimed share A.[30] The results for the stakes dummy variables indicate that contestants are less likely to make hardball announcements if the jackpot is relatively small. The raw frequencies also demonstrate this pattern. In the bottom jackpot quartile, where the stakes are €15,195 or lower, only 13 percent of the contestants who

[30] Only one of the contestants who claimed share B made a hardball announcement. Including contestants who claimed share B does not influence our results.

claim share A accompany their claim with a hardball announcement. For the second, third and fourth quartile, this percentage is 32, 25 and 39 percent, respectively. Furthermore, the results for *Variance shares* indicate that hardball announcements are more likely if the differences between the shares are larger. If the share distribution is 60/30/10, roughly 17 percent of the contestants make a hardball announcement. For the more extreme 65/25/10 and 70/20/10 divisions, the proportions of hardball announcements are 24 and 48 percent, respectively. Demographic and contribution variables do not influence the likelihood that a contestant makes a hardball announcement.

We now turn to contestants' actual concession behavior. Models 3 through 6 display the results from the probit regression analyses on contestants' decisions to lower their claim (1) or not (0). Model 3.3 and Model 3.4 are restricted to variables that are exogenous to the bargaining stage, while Model 3.5 and Model 3.6 also include initial claims and hardball announcements. The analysis is performed for the subset of contestants who initially claimed share A or B and did not reach agreement immediately, as only these contestants have the opportunity to lower their claim.

The coefficients for the stakes dummies suggest a U-shaped effect of the jackpot size. For observations from the second and third quartile, the likelihood of a concession is significantly smaller than for observations from the first. In the first and fourth quartile, concessions are approximately equally likely. The proportions of contestants lowering their claims in the four quartiles are 63%, 43%, 41% and 55% respectively.

The way the jackpot is divided across the three shares A, B, and C has no significant influence here. The demographic variables are also insignificant. For the effect of contributions, however, we find an interesting asymmetry. While the initial claims are especially driven by positive contributions, concessions turn out to be mostly determined by negative contributions. The *Contribution overall* variable does not reach significance, but when we distinguish between positive and negative contributions, we find that *Contribution incorrect* is significant. In terms of marginal effects, a 10 percentage-point increase in *Contribution incorrect* implies a 7.2 percentage-point increase in the likelihood of a concession.

Table 3.4: Probit Regression Results on Hardball Announcements and Concessions

The table displays results from the probit regression analyses on contestants' hardball announcements (Model 3.1 and 2) and concessions (Model 3.3, 4, 5 and 6) in the bargaining stage of the British TV game show *Divided*. The hardball (concession) analyses are performed on the subset of contestants who initially claimed share A (who initially claimed share A or B and did not reach agreement immediately). Definitions of variables are as in the previous tables. Standard errors are corrected for clustering at the team level, *p*-values are in parentheses.

	Hardball announcements		**Concessions**			
	Model 3.1	Model 3.2	Model 3.3	Model 3.4	Model 3.5	Model 3.6
Demographic characteristics						
Age	-0.008 (0.553)	-0.007 (0.625)	0.000 (0.998)	-0.007 (0.562)	-0.003 (0.798)	-0.009 (0.433)
Gender (male=1)	-0.161 (0.576)	-0.123 (0.696)	0.080 (0.773)	-0.096 (0.715)	0.005 (0.987)	-0.184 (0.522)
Education (high=1)	0.011 (0.972)	0.033 (0.920)	-0.003 (0.992)	-0.057 (0.852)	0.064 (0.847)	0.034 (0.920)
Situational variables						
First mover (first=1)	-0.106 (0.701)	-0.121 (0.664)	0.179 (0.559)	0.300 (0.359)	0.154 (0.623)	0.278 (0.394)
Stakes 2nd quartile	0.942 (0.040)	0.934 (0.044)	-0.529 (0.029)	-0.532 (0.032)	-0.503 (0.056)	-0.535 (0.046)
Stakes 3rd quartile	0.767 (0.092)	0.746 (0.106)	-0.583 (0.009)	-0.545 (0.018)	-0.541 (0.013)	-0.508 (0.022)
Stakes 4th quartile	1.032 (0.015)	1.009 (0.021)	-0.239 (0.302)	-0.156 (0.514)	-0.175 (0.479)	-0.090 (0.720)
Variance shares	30.679 (0.033)	31.002 (0.032)	-6.632 (0.524)	-9.183 (0.385)	-4.396 (0.639)	-7.050 (0.461)
Contribution variables						
Contribution overall	-0.006 (0.996)		-1.273 (0.273)		-1.620 (0.201)	
Contribution correct		-0.533 (0.772)		0.085 (0.961)		-0.539 (0.764)
Contribution incorrect		-0.005 (0.994)		1.801 (0.035)		2.001 (0.027)
Claim variables						
Initial claim A (A=1)					0.388 (0.239)	0.524 (0.143)
Announce hardball (hardball=1)					-0.985 (0.002)	-1.000 (0.003)
Opp. announce hardball (hardball=1)					0.491 (0.098)	0.499 (0.099)
Constant	-2.696 (0.008)	-2.538 (0.029)	0.993 (0.151)	0.097 (0.914)	0.701 (0.316)	-0.273 (0.750)
Log pseudo-likelihood	-53.84	-53.79	-76.59	-74.24	-69.87	-67.48
McFadden R^2	0.102	0.103	0.039	0.069	0.123	0.153
N	102	102	115	115	115	115
Number of teams	43	43	39	39	39	39

Model 3.5 and Model 3.6 also incorporate the effects of initial claims and hardball announcements. Contestants who make hardball announcements turn out to put their money where their mouth is. A hardball announcement indeed decreases the likelihood of making a concession. A player who says that she will not budge is approximately 37 percentage points less likely to make a concession. Hardball announcements are also considered as credible threats: when contestants face an opponent who made a hardball announcement, they are approximately 20 percentage points more likely to make a concession. Further, concessions are equally likely for players who initially picked the top prize and for those who picked the middle prize.

Final Payoffs

A contestant's bargaining outcome can be defined in two different ways: relative to others and relative to the initial size of the jackpot. First, we consider payoffs relative to those of the opponents. That is, we look at the share (A, B or C) that a contestant ends up with. Players who fail to reach agreement and go home empty-handed are excluded from this analysis. Note that the stakes and the differences between the three percentage shares cannot influence the likelihood of receiving a particular share, given the fact that all shares are awarded and that these factors are constant at the team level.

Table 3.5 displays the results. Model 3.1 and Model 3.2 are restricted to variables that are exogenous to the bargaining stage, while Model 3.3 and Model 3.4 also include initial claims and hardball announcements. In line with previous analyses, the restricted models show that those who contributed more to the jackpot are more likely to end up with a larger share. This effect is driven by both positive and negative contributions. The extended models similarly demonstrate the relation between contributions and bargaining outcomes. Because the effect of positive contributions on the bargaining process is reflected in the initial claims, including initial claims in the model reduces the significance of the measure for positive contributions. As before, we find no effect of demographic characteristics and there is no first-mover advantage.

Table 3.5:Ordered Probit Regression Results on Share Won

The table displays results from the ordered probit regression analyses on contestants' final claims A (3), B (2) or C (1) when agreement is reached in the bargaining stage of the British TV game show *Divided*. Definitions of variables are as in the previous tables. Standard errors are corrected for clustering at the team level, *p*-values are in parentheses.

	Model 3.1	Model 3.2	Model 3.3	Model 3.4
Demographic characteristics				
Age	-0.011 (0.357)	-0.010 (0.386)	0.000 (0.979)	0.002 (0.887)
Gender (male=1)	-0.011 (0.956)	0.066 (0.748)	0.138 (0.521)	0.245 (0.275)
Education (high=1)	-0.020 (0.925)	-0.003 (0.989)	-0.146 (0.590)	-0.120 (0.659)
Situational variables				
First mover (first=1)	-0.132 (0.687)	-0.181 (0.590)	-0.115 (0.735)	-0.160 (0.639)
Contribution variables				
Contribution overall	2.871 (0.002)		2.300 (0.030)	
Contribution correct		2.969 (0.037)		2.216 (0.167)
Contribution incorrect		-1.260 (0.049)		-1.243 (0.043)
Claim variables				
Initial claim A (A=1)			6.302 (0.000)	5.927 (0.000)
Initial claim B (B=1)			5.780 (0.000)	5.511 (0.000)
Announce hardball (hardball=1)			0.889 (0.011)	0.869 (0.017)
Opp. announce hardball (hardball=1)			-0.522 (0.018)	-0.547 (0.014)
α_1	0.438 (0.186)	0.069 (0.910)	6.288 (0.000)	5.504 (0.000)
α_2	1.357 (0.000)	0.996 (0.107)	7.393 (0.000)	6.622 (0.000)
Log pseudo-likelihood	-110.10	-109.52	-94.37	-93.70
McFadden R^2	0.046	0.051	0.182	0.188
N	105	105	105	105
Number of teams	35	35	35	35

The results for Model 3.3 and 4 show that contestants who announce to play hardball fare better than others. In terms of marginal effects, those who announce not to back down are approximately 30 percentage points more likely to go home with the top share and 29 percentage points less likely to go home with the bottom share. At the same time, if an opponent makes a hardball announcement, this has a negative effect on a contestant's final share. More specifically, it decreases the likelihood of receiving the top share by 15 percentage points, and increases the likelihood of ending up with the bottom share by 21 percentage points.

These analyses of contestants' final claims ignore the efficiency of the bargaining process. A contestant may feel like a winner if she secured more money than her opponent after fighting for share A for 70 seconds, but if she could have won more money in an absolute sense by directly going for share B, then objectively the latter approach would have been the better strategy (ex-post). Next, we therefore analyze the money that players take home as a fraction of the initial jackpot.

Table 3.6 shows the results, again for two models with exogenous variables only and for two extended models with variables for initial claims and hardball announcements. As in the previous analyses, those who contributed more to the jackpot secure a larger payoff. Again the effect is non-trivial: a 10 percentage-point increase in contribution increases earnings by approximately 4 percent of the initial jackpot.

Hardball announcements clearly frustrate the bargaining process. The previous analysis showed that contestants who announce hardball do well relative to others, but the present one points out that these players do not go home with larger amounts in an absolute sense. Their opponents, however, are significantly worse off from both a relative and an absolute perspective.

The stakes dummy variables point at a U-shaped effect of the initial jackpot: contestants generally fare better at the top and bottom stakes quartiles and less so in between. This result is in line with the U-shaped pattern for concession behavior: concessions occur more often at the top and bottom stakes quartiles, which shortens the time to resolution for these quartiles. In addition, contestants are less likely to reach a speedy consensus if the differences between the prizes are larger.

This final analysis is the sole analysis that generates a significant effect for a demographic variable: older contestants take home a smaller part of the initial size of the pie. Again, there is no evidence of a first-mover advantage.

Table 3.6: OLS Regression Results on Prize Won / Initial Jackpot

The table displays results from the OLS regression analyses on the fraction of the initial jackpot that the contestant takes home in the British TV game show *Divided*. Definitions of variables are as in the previous tables. Standard errors are corrected for clustering at the team level, *p*-values are in parentheses.

	Model 3.1		Model 3.2		Model 3.3		Model 3.4	
Demographic characteristics								
Age	-0.003	(0.022)	-0.003	(0.032)	-0.002	(0.030)	-0.002	(0.041)
Gender (male=1)	-0.019	(0.484)	-0.017	(0.551)	-0.014	(0.614)	-0.012	(0.682)
Education (high=1)	-0.017	(0.620)	-0.017	(0.641)	-0.025	(0.496)	-0.025	(0.495)
Situational variables								
First mover (first=1)	-0.004	(0.899)	-0.009	(0.809)	0.002	(0.951)	-0.001	(0.963)
Stakes 2nd quartile	-0.089	(0.056)	-0.089	(0.057)	-0.083	(0.124)	-0.082	(0.127)
Stakes 3rd quartile	-0.082	(0.063)	-0.083	(0.061)	-0.079	(0.126)	-0.079	(0.123)
Stakes 4th quartile	0.020	(0.655)	0.019	(0.671)	0.033	(0.531)	0.032	(0.543)
Variance shares	-4.037	(0.009)	-4.002	(0.011)	-3.769	(0.032)	-3.726	(0.036)
Contribution variables								
Contribution overall	0.388	(0.000)			0.374	(0.001)		
Contribution correct			0.437	(0.007)			0.436	(0.003)
Contribution incorrect			-0.114	(0.092)			-0.115	(0.086)
Claim variables								
Initial claim A (A=1)					0.071	(0.059)	0.079	(0.049)
Initial claim B (B=1)					0.072	(0.091)	0.089	(0.068)
Announce hardball (hardball=1)					0.053	(0.175)	0.054	(0.170)
Opp. announce hardball (hardball=1)					-0.068	(0.050)	-0.070	(0.050)
Constant	0.285	(0.000)	0.306	(0.001)	0.215	(0.007)	0.223	(0.009)
Adjusted R^2	0.123		0.117		0.152		0.148	
N	129		129		129		129	
Number of teams	43		43		43		43	

3.5 Conclusions and Discussion

The British TV game show *Divided* offers the opportunity to study high stakes bargaining in a controlled setting and for a diverse subject pool. In line with equity theory (Homans, 1958; Adams, 1965; Walster, Berscheid and Walster, 1973), we find strong evidence that contestants derive entitlements from their contributions to the jackpot. Interestingly, positive and negative contributions appear to have different effects: positive contributions drive contestants' opening claims, while negative contributions are important for concessions during the subsequent free-form negotiation process. One explanation for this asymmetry is that those with negative contributions initially consider such contributions to be innocent mistakes for which they should not be held accountable, and that subsequent communication works to promote a more objective, less self-biased view. The asymmetry is, however, also in line with query theory (Johnson, Häubl and Keinan, 2007): contestants' initial focus on positive contributions occurs when the problem is framed in positive terms ("what share do you deserve?"), but switches to negative contributions when the framing becomes negative ("who should move their claim downward?").

Announcing a hardball strategy of not backing down turns out to not be beneficial. Contestants who used this threat do well relative to others, but they do not manage to get larger amounts in an absolute sense. Their opponents are worse off, because contestants who make a hardball announcement also walk the walk: they are less likely to make a concession, and thus frustrate the team's chances of coming to resolution. The inefficacy of a hardball approach is in line with game-theoretic reasoning, as cheap talk should not yield any advantage if actors' interests are not sufficiently aligned (Crawford and Sobel, 1982) and neither should simple strategies that anyone can follow.

We find that contestants are more likely to announce a hardball strategy if the stakes are higher. Actual concession behavior appears to be affected by the stakes in a non-monotonic fashion. The likelihood of concessions in our sample is U-shaped: contestants display a higher propensity to make concessions at the lowest and highest stakes quartiles, and less so in between. As a result, consensus is more easily reached if the jackpot is at the lower or upper quartile, and contestants retain a larger share of the jackpot in these cases.

We find little to no effect of background characteristics on bargaining behavior and outcomes. This is in line with earlier results that demographic characteristics hardly explain any variance in bargaining performance (Rubin and Brown, 1975; Thompson, 1990; Walters, Stuhlmacher and Meyer, 1998; Stuhlmacher and Walters, 1999; Elfenbein *et al.,* 2008), except in situations that have substantial structural ambiguity (Bowles, Babcock and McGinn, 2005). Furthermore, there seems to be no first-mover advantage. Those who get to make their claim early do not earn more. This suggests that first-mover advantages might be restricted to bargaining games of alternating offers.

We conclude with some thoughts on the generalizability of our results. There are three possible grounds for external validity concerns: the specific bargaining game, possible selection effects, and the unusual decision environment.

First, the bargaining game indeed has some notable characteristics. Presumably, it was designed to make reaching consensus relatively difficult in order to promote the entertainment value of the show. Four characteristics spring to mind: the unanimity rule for reaching a decision, the use of face-to-face free-form communication, the indivisibility of the shares to be divided, and the continuous costs to bargaining. These aspects will undoubtedly have their effect on the bargaining process.

The unanimity rule most likely makes it more difficult to reach consensus. Miller and Vanberg (2013) conducted an experiment with a highly structured bargaining game in which three subjects had to agree on the division of a joint pie by either the unanimity rule or the majority rule, and find that the effect of requiring unanimity in itself is aggravated by an increased tendency among subjects to reject offers. Communication has been found to affect bargaining in several ways. Bazerman *et al.* (2000) review the negotiation literature and conclude that face-to-face communication has the potential to foster agreements by development of rapport, decrease misunderstandings, and increase truth telling. At the same time, they find that face-to-face communication can lead to more pressure tactics and impasses if tensions are high.

We are unaware of studies that examine the impact of indivisible shares and continuous costs to bargaining. For indivisible shares, the effect is probably highly dependent on the sizes of the differences between the shares. Dividing three equal shares will lead to immediate consensus under all circumstances. The highly unequal shares used in *Divided*, however, are likely to hamper agreement in comparison with less restrictive division rules. Continuous costs to bargaining probably increase tension. More discrete rules, such as a drop of 10 percentage points every 10 seconds or a once-and-for-all destruction of the jackpot after 100 seconds with no destruction in between, would have given subjects more time to negotiate before cost had to be incurred. It also seems reasonable to assume that the bargaining costs lead to some kind of sunk-cost effect, making players less willing to back down from their initial claim. Overall, these four characteristics of our specific bargaining game thus appear to make consensus formation relatively difficult. On the other hand, aspects such as the large incentives (as compared to standard experiments) and the fact that the team has cooperated before playing the game may work to promote or facilitate consensus formation.

The main focus of our analyses has not been on the general level of consensus. Rather, we focused on comparative statics, and it is unclear to what extent the specific bargaining game influences the effects of variables such as contributions and stakes. Communication, for example, might increase the importance of contributions by allowing subjects to reach consensus regarding who has earned which share, but it can also hamper the role of contributions because it allows for more pressure tactics. While such sensitivities are an interesting domain for further research, face-to-face communication, continuous cost to bargaining, and the requirement of a unanimous decision are features that our game show environment shares with many real world settings.

The second generalizability concern relates to possible selection effects. Contestants self-select into auditions and are then selected by producers of the show to play the game for real. It is unclear to which degree such selection processes may have influenced our findings. Still, our sample varied widely in terms of background characteristics, seemingly forming a cross-section of middle-class society. It is much closer to a cross-section of the general population than university students commonly employed in conventional laboratory experiments. Furthermore, take note that selection procedures are not unique to game shows, and form an intrinsic part of almost any field or laboratory setting.

Last, a game show setting may impact behavior. While there is no live studio audience, contestants know that many people will observe their behavior on TV. This makes that the bargaining game is not strictly one-shot, as contestants' behavior and outcomes might affect their reputation. The specific setting provides an incentive to fight harder, as one may not want to appear weak on TV. However, being stubborn and then consequently loosing (a large fraction of) the jackpot is also an outcome to be avoided. Relatedly, the game show format might create a desire to "win the contest" and go home with more money than fellow team members, but contestants may also interpret the "contest" as a challenge to come to resolution with the people they teamed up with.

We do not consider the possible influences of the specific decision environment to render our findings less interesting or less predictive of behavior in other settings. In laboratory and real-life situations there is normally always some degree or form of scrutiny, and each specific setting and contextual aspect will cause particular motives to be more prominent than others. For example, people normally approach bargaining for a pay increase at work differently from bargaining with a stranger over a second-hand car, and differently from bargaining with a spouse over who should do the ironing. That is not to say that we cannot learn anything more general from observing behavior in specific settings. We cannot study behavior under each and every possible set of conditions, and the optimal approach to assess the robustness and generalizability of findings is therefore to study behavior in a limited number of diverging settings. The contribution of the present chapter should be evaluated in this light. We employed the unique features of a TV game show to study bargaining behavior outside the laboratory and for stakes that are impossible to replicate in a behavioral laboratory.

One of our main results is that entitlements derived from contributions are an important driving force behind bargaining behavior and outcomes. The importance of moral property rights in our high-stakes environment refutes the commonly held view that fairness concerns are unimportant when monetary incentives are sufficiently large. Another main result is perhaps in the inefficacy of adopting a hardball strategy to obtain a bigger share of the pie. Due to bargaining costs, the total pie in our game shrinks such that there is no advantage left for the threatening party and others are worse off. The recent political impasse in the US may serve as an illustration of the broader relevance of this finding. In a very different context than our show, pie-decreasing hardball moves of opposing politicians led to a costly shut down of the government and unprecedented threats to default on the government's debt.

Chapter 4 | Risky Choice in the Limelight

This chapter examines how risk behavior in the limelight differs from that in anonymity. In two separate experiments we find that subjects are more risk averse in the limelight. However, risky choices are similarly path dependent in the different treatments. Under both limelight and anonymous laboratory conditions, a simple prospect theory model with a path-dependent reference point provides a better explanation for subjects' behavior than a flexible specification of expected utility theory. Additionally, our findings suggest that ambiguity aversion depends on being in the limelight, that passive experience has little effect on risk taking, and that reference points are determined by imperfectly updated expectations.

This chapter is based on the paper "Risky Choice in the Limelight", co-authored by Guido Baltussen and Martijn J. van den Assem (Baltussen, van den Assem, and van Dolder, 2013). We are grateful to Han Bleichrodt, Chen Li and Peter Wakker for their constructive and valuable comments. We thank Thierry Post for his help in designing and conducting the experiments, Marc Schauten for acting as game show host, and Nick de Heer for his skillful research assistance. The chapter benefited from discussions with seminar participants at the Erasmus University of Rotterdam, and with participants of BDRM 2012 Boulder, FUR 2012 Atlanta, M-BEES 2012 Maastricht, SABE 2012 Granada, TIBER 2012 Tilburg, and SPUDM 2013 Barcelona.

4.1 Introduction

Individual decision-making is at the core of both economics and psychology, and continuous research efforts have resulted in a rich literature. Still, a persistent concern about empirical research in these fields is that specific contextual aspects may restrict the generalizability of findings. Each laboratory or field setting provides its own unique context that cannot be disregarded a priori (Loewenstein, 1999; Levitt and List, 2007a, 2007b; Falk and Heckman, 2009; Camerer, 2011). One particular aspect of the context is the degree of public scrutiny under which a decision is made. Psychological research in this area indicates that the mere presence of others can facilitate performance in simple tasks but impair it in more complex ones (Zajonc, 1965; Bond and Titus, 1983), and that the expectation that one may have to justify one's decisions to observers creates a desire to make decisions that others will judge favorably (Lerner and Tetlock, 1999).

Economists have demonstrated relatively little interest in context effects, and their studies on public scrutiny have primarily focused on social preferences (Levitt and List, 2007a, 2007b). Surprisingly, whether and how public scrutiny influences risky choice has received relatively little attention from both economists and psychologists.[31] In our professional and private lives we make decisions under varying degrees of scrutiny, and mapping the influence of this contextual aspect is therefore an important step in further broadening the scope of our understanding of risky choice. Also, from a methodological point of view, it is useful to know to what extent findings on risk preferences from a behavioral laboratory generalize to real world situations with more scrutiny, and whether risky choices observed in a high-scrutiny field setting resemble those in a situation with more privacy.

[31] Weigold and Schlenker (1991) find evidence that subjects display a degree of risk tolerance they believe to be judged favorably by observers. Vieider (2009) finds that loss aversion decreases when subjects are made accountable. He attributes this finding to the ease with which his subjects could recognize loss aversion as a bias and their wish to avoid negative judgments that would result from displaying this bias. Neither of these two studies used real incentives, which makes it costless for subjects to make a choice that is not truly preferred but thought to be more justifiable in the eyes of onlookers. For hypothetical and incentivized tasks, Miller and Fagley (1991), Takemura (1993, 1994), and Vieider (2011) find that the effect of framing outcomes as gains or losses decreases when subjects are made accountable.

A special example of the relevance of our research question is in the growing literature that studies decision making under risk on the basis of television (TV) game shows. Shows that have been used include *Card Sharks* (Gertner, 1993), *Jeopardy!* (Metrick, 1995), *Illinois Instant Riches* (Hersch and McDougall, 1997), *Lingo* (Beetsma and Schotman, 2001), *Hoosier Millionaire* (Fullenkamp, Tenorio and Battalio, 2003), *Who Wants to Be a Millionaire?* (Hartley, Lanot and Walker, 2006), and *Deal or No Deal* (Post *et al.,* 2008). These shows offer unique opportunities to increase our understanding of how individuals and households make significant risky decisions such as stock market investments and the purchase, insurance, and financing of property.[32] The large prizes in game shows enable researchers to study behavior for stakes that are considerably more consequential than those typically employed in experiments, and the simple and well-defined decision problems impose fewer auxiliary assumptions than uncontrolled field data. Also, even though selection effects are inevitable, game show contestants generally resemble a cross-section of the general population more closely than subjects in most conventional experiments. Given the attractive and distinguishing combination of features that game shows can have, more game show-based papers are likely to appear. Some critics, however, question the external validity of game show research, arguing that contestants' choices might be influenced by pressures from the audience and distress from being in the limelight. As noted by Gertner (1993, p.519), for example: "if contestants care about the entertainment they provide, they may make riskier decisions than they otherwise would."

First and foremost, the present chapter contributes to the risky choice literature by comparing risky decision making in and out of the limelight. Next to this, it also adds some evidence to the literature on ambiguity aversion by comparing the effect of ambiguity under these two conditions, and to a recent literature on the effect of experience on choices by comparing the behavior of subjects with and without passive

[32] Game shows have been deployed on various other research domains as well, including strategic decision making (Bennett and Hickman, 1993; Berk, Hughson and Vandezande, 1996; Tenorio and Cason, 2002), discrimination (Levitt, 2004; Antonovics, Arcidiacono and Walsh, 2005) and cooperative behavior (List, 2004a, 2006; Belot, Bhaskar and van de Ven, 2010a; Oberholzer-Gee, Waldfogel and White, 2010; van den Assem, van Dolder and Thaler, 2012).

experience. Finally, our estimations of structural models of choice add to the literature on reference point formation.

To analyze how risky choice in the limelight differs from that under standard experimental conditions, we conducted two incentivized experiments that mimicked the game of the TV show *Deal or No Deal* (henceforth: DOND). The next section describes DOND and explains why we used this particular game. In both experiments, we employed laboratory and limelight treatments. In the laboratory treatments, subjects made decisions in the anonymity of a standard, computerized laboratory setting as typically employed in economic experiments. In the limelight treatments, subjects made their choices in a simulated game show environment, which included a live audience, a game show host, and video cameras.

By using a game show environment to create public scrutiny, we also shed light on the validity of game shows as natural risky choice experiments. For domains other than risky choice, a number of studies have investigated this specific issue before. Tenorio and Cason (2002), Healy and Noussair (2004), and Antonovics, Arcidiacono and Walsh (2009) observe how students play *The Price is Right* and *The Weakest Link* under laboratory conditions and find that their behavior or performance is similar to that of contestants in the TV show.

We consider two ways in which the differences between the treatments can influence risky choice. First, we investigate whether the general degree of risk taking differs between treatments. Second, we examine whether the pattern of path-dependent risk behavior is different. Earlier DOND-based research has found that people show path dependency in the sense that they take more risk if the game develops either substantially worse or substantially better than earlier expectations (Post *et al.,* 2008). These two effects are known as the break-even and house-money effect (Kameda and Davis, 1990; Thaler and Johnson, 1990).

If only the general degree of risk taking is affected, this is problematic only in so far as risk preferences are measured in one setting and used to derive point predictions about behavior in another setting. It would imply that it is inappropriate to apply the same risk preference *parameters* across different settings. If, however, the pattern in risky choice is different, the repercussions are potentially more involved, because it would mean that we cannot use the same type of risky choice *model* across different settings.

Our results show that subjects are more risk averse in the limelight than in the anonymity of a typical behavioral laboratory. Simple statistics, probit analyses, and structural choice model estimations consistently lead to this conclusion for both our experiments. The estimates for structural choice models suggest that the impact of the limelight on risk preference parameters is substantial.

At the same time, however, we find a similar pattern of path-dependent risk behavior in the laboratory and limelight treatments. Under both experimental conditions, our simple prospect theory-inspired model (Kahneman and Tversky, 1979; Tversky and Kahneman, 1992) with a path-dependent reference point provides a better explanation for subjects' behavior than a flexible specification of expected utility theory. Although our study is not designed to point out whether prospect theory or expected utility theory has greater descriptive power and any conclusion in this direction would depend on the precise empirical implementation of these two theories, it does show that the combination of elements included in our prospect theory model comes closer to the appropriate descriptive model of risky choice, and that this finding holds both in and out of the limelight.

Three other noteworthy findings from our analyses are related to ambiguity aversion, the effect of experience, and reference-point formation. First, a design difference between the two sets of experiments that we conducted reveals that the effect of ambiguity depends on being in the limelight or not. Under limelight conditions, subjects take less risk in tasks where they experience some uncertainty about the distribution of possible outcomes than in tasks where the distribution is known. This difference in behavior is absent under laboratory conditions. Second, passive experience does not seem to affect loss aversion or risk aversion in general. One of our experiments featured a comeback treatment with subjects who had seen others perform the experimental task at an earlier occasion. Comparisons between treatments show that their behavior is largely similar to that of inexperienced subjects. Last, we find evidence that preferences are based on imperfectly updated expectations. For all treatments, the parameter estimates of our prospect theory model indicate that subjects' reference points are influenced by their initial beliefs about task outcomes.

The chapter proceeds as follows. Section 4.2 describes the design, procedure and results of our first experiment. Section 4.3 reports on our second experiment. Section 4.4 discusses our results and concludes.

4.2 First Experiment

Design and Procedure

The experiment followed the basic setup of the popular TV game show *Deal or No Deal*. In DOND, contestants are repeatedly asked to make choices between a sure amount and a risky lottery. At the start, a contestant chooses one case out of a total of 26 numbered (brief)cases. Each closed case contains one of the game's 26 randomly distributed and widely ranging monetary amounts. After selecting this personal case, a contestant has to select six of the other cases to be opened. The prizes in these cases are revealed and are no longer in play, thereby increasing the information on the prize in the contestant's personal case. After the contents of six cases have been revealed, an imaginary "banker" offers to buy the contestant's case. If the contestant decides "Deal", she receives the amount offered and the game ends. If the contestant decides "No Deal", the game continues and she has to open five additional cases. Based on the then remaining set of 15 prizes, the banker makes a new offer. The contestant again has to decide either "Deal" or "No Deal". After a "No Deal", this process continues either until the contestant accepts an offer, or until no case other than the contestant's own case is left and she receives the content of this case. The game lasts for a maximum of nine rounds. The number of cases to be opened in each round is 6, 5, 4, 3, 2, 1, 1, 1, and 1, reducing the number of remaining cases from 26 to 20, 15, 11, 8, 6, 5, 4, 3, 2, and finally 1. Figure 4.1 presents a schematic overview of the course of the game.

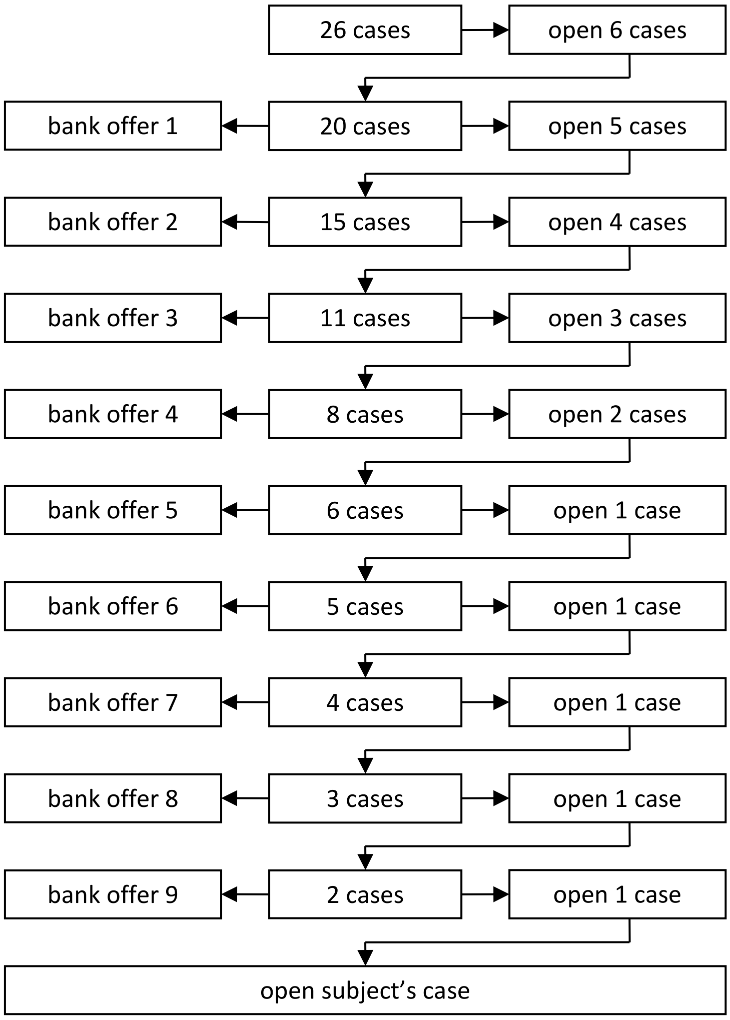


Figure 4.1: Flow Chart of the Game

In each of a maximum of nine game rounds, the subject chooses a given number of cases to be opened. After the prizes in the chosen cases are revealed, an imaginary banker offers to buy the subject's own case. If the subject accepts the offer ("Deal"), she receives the amount offered and the game ends. If the subject rejects the offer ("No Deal"), play continues and she enters the next round. If the subject decides "No Deal" in the ninth round, she receives the prize in her own case. (Taken from Post *et al.,* 2008.)

In the experiment, subjects played DOND for real incentives in either a computer laboratory (laboratory treatment) or in a classroom mimicking a TV studio (limelight treatment). The prizes in the experiment were equal to the prizes used in the original Dutch edition of the TV show, scaled down by a factor of 10,000. The smallest amounts were rounded up to one cent. The resulting set of prizes was €0.01 (9 times due to rounding up); €0.05; €0.10; €0.25; €0.50; €0.75; €1; €2.50; €5; €7.50; €10; €20; €30; €40; €50; €100; €250; €500. The distribution of the prizes was clearly positively skewed, with a median of €0.63 and a mean of €39.14. Figure 4.2 demonstrates how the game was shown to subjects.

The laboratory treatment was conducted as a typical economic experiment. Subjects played DOND in the quiet, controlled environment of a computerized laboratory, and made their choices on a private computer terminal. The setting was designed to minimize potential scrutiny from other subjects. In particular, computers surrounding a given subject were empty and a sunken screen and dividers were used to ensure privacy.

The limelight treatment was designed to replicate a TV studio as closely as possible. The experiment took place in a theater-style lecture room. Subjects made their decisions on a lightened stage in front of a live audience, consisting of fellow students and some university employees. They were guided through the experiment by a game show host, played by a popular lecturer. Furthermore, video cameras were pointed at the subject on stage. The game was shown on a computer monitor in front of the subject and projected on a large screen for the audience. Members of the audience were allowed to applaud, shout hints, and the like. Before a game started, the host had a brief introductory talk with the subject on stage, covering basic topics such as the subject’s name, age, favorite sports, and other interests.

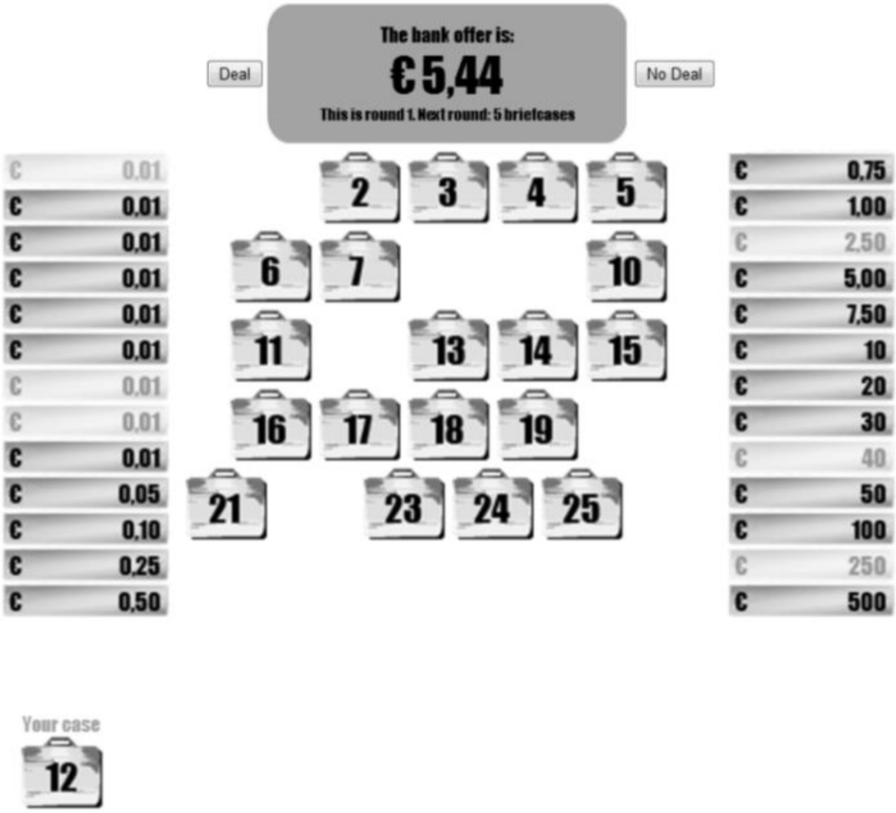


Figure 4.2: Example of the Game as Displayed on the Experimental Screens

The various prizes are listed in the columns on the left and right side. Prizes that are eliminated are blurred. The current bank offer is shown at the top, and the subject or host can select either “Deal” or “No Deal” by clicking on the respective button. The remaining cases are shown in the center of the screen, while the subject’s own case is in the bottom left-hand part. This example shows the two options open to a subject after opening six cases in the first round: accept a bank offer of €5.44, or continue to play with the remaining 20 cases. Note that a comma is used to separate decimals here, as this is common for our subjects.

The data from the limelight treatment has previously been analyzed in Post *et al.* (2008).[33] To facilitate comparisons with the actual game show in that study, each subject replayed one of the first 40 scenarios from the Dutch version of DOND: independent of the order in which a subject opened the numbered cases, the order in which the prizes and the offers appeared corresponded exactly to the original scenario. In addition, we matched the gender of subjects and TV contestants: female (male) subjects were randomly assigned to scenarios from female (male) contestants. We did not select these 40 scenarios to encourage or avoid particular situations or behaviors. In fact, subjects played a randomly chosen game that had been generated by chance at an earlier point in time.[34] The instructions were as similar as possible to those that had been handed out to TV show contestants. Subjects received the original Dutch instructions used for the TV version, plus a cover sheet explaining the experiment. We did not impose any time constraint.

Subjects were randomly selected from a larger population of business or economics students at the Erasmus University of Rotterdam who had applied to participate in economic experiments. In total, 40 subjects took part in the limelight treatment, and 40 took part in the laboratory treatment: one for each of the 40 scenarios in both treatments. We subdivided subjects in the limelight treatment across two separate sessions. In total, 80 students were invited to the two limelight sessions, approximately 40 per session. This was done to ensure a sufficiently large audience and to create a buffer in case some subjects would not show up. After one subject had finished playing the game, a new subject was selected to play, until 20 subjects had played the game. Hence, approximately half of the students in each session were selected to play. Subjects were paid according to the outcome of their game. Subjects who were not selected received no pay. Each game lasted about five to ten minutes, and an entire session lasted

[33] The limelight treatment was employed there to analyze the isolated effect of the amounts at stake. (Another treatment was conducted under identical limelight conditions but used stakes that were a factor of ten larger.)

[34] If a subject played on longer than the original contestant, we had no information on eliminated prizes and bank offers from that point onward. We then randomly selected the eliminated prizes ourselves (holding them constant across treatments) and set the offers according to the pattern observed for the TV episodes.

approximately 2.5 hours. The 40 subjects who were selected for the laboratory treatment were similarly subdivided across two different sessions. In each session, 20 subjects played the game simultaneously.

Using the game of DOND has several benefits. Its appealing qualities have attracted considerable research attention, making it the most frequently studied game show in the domain of risky choice (Blavatskyy and Pogrebna, 2010; Brooks *et al.,* 2009a, 2009b; Deck, Lee and Reyes, 2008; Post *et al.,* 2008). The game involves only simple stop-go decisions ("Deal" or "No Deal") that require no or minimal skill, knowledge or strategy. Moreover, the dynamic nature of the game allows to not only compare general levels of risk taking between treatments, but also the pattern of path dependence. In addition, subjects may find it relatively natural to make decisions in front of an audience when the task at hand is from a TV game show, and the entertainment value of DOND may help to involve the audience in the game. The great popularity of the game on TV brings the advantage that it is generally well understood by subjects.

Descriptive Statistics and Probit Analysis

In total, we observed 579 decisions made by 80 subjects. A crude way to investigate differences in risky choice between the treatments is to compare subjects' stop rounds. A stop round is the round in which a subject decides to accept the bank offer ("Deal"), or 10 if she rejects all nine offers. The bank offer typically starts as a small percentage of the average remaining prize, and gradually increases as the game proceeds. Deciding "Deal" at a relatively early (late) stage thus implies a relatively high (low) degree of risk aversion.

Figure 4.3 shows the distribution of the stop round for both treatments. Subjects in the limelight treatment decide to "Deal" earlier than subjects in the laboratory treatment. The average stop round in the limelight treatment is 6.93, compared to 7.93 in the laboratory treatment. The difference of exactly one round is statistically significant (t-test: p = 0.019; Mann-Whitney U test: p = 0.021).

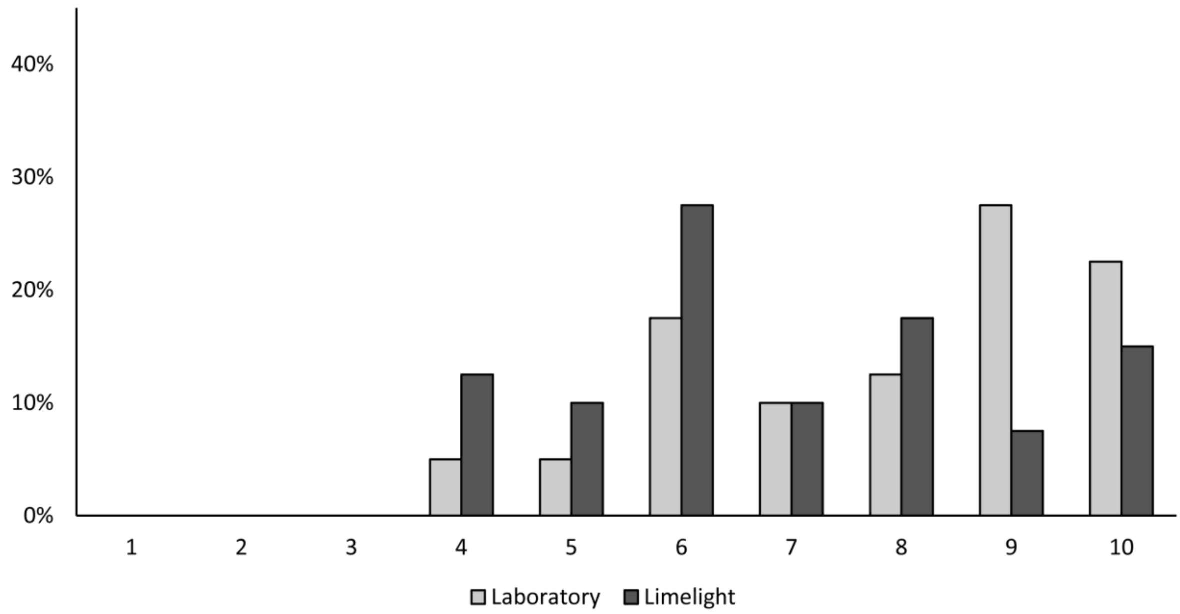


Figure 4.3: Distributions of Stop Rounds (First Experiment)

The figure depicts the distribution of stop rounds for the two treatments of our first experiment. The stop round is the round in which the bank offer is accepted (“Deal”), or 10 for subjects who rejected all offers. In the laboratory treatment, subjects played the game in a standard economic laboratory setting, while in the limelight treatment subjects played the game in an environment mimicking a TV studio with live audience.

The stop round is a crude measure as it does not reflect differences in the actual bank offer, the stakes, or the risk of continuing play. To control for these factors, we perform a probit regression analysis. The dependent variable is the subject’s decision, taking the value of 1 for “Deal” and 0 for “No Deal”. We explain subjects’ decisions using the following variables:

- *EV*/100: included to control for the stakes, and calculated as the current average remaining prize (divided by 100 Euros for more convenient regression coefficients);

Table 4.1: Probit Regression Results (First Experiment)

The table displays the maximum likelihood estimation results of a probit model aimed at explaining the decisions of the subjects in the laboratory (*N* = 40) and limelight (*N* = 40) treatment of our first experiment. The dependent variable is the subject's decision, with a value of 1 for "Deal" and 0 for "No Deal". *EV* is the current average remaining prize in Euros, *BO* is the bank offer in Euros, *Stdev* is the standard deviation of the distribution of the average remaining prize in the next game round, and *Limelight* is a dummy variable that takes a value of 1 for observations from the limelight treatment. In addition to the maximum likelihood estimates for the regression coefficients, the table reports the log-likelihood (LL), McFadden R^2, and the total number of observations (No. obs.). The *p*-values (within parentheses) are corrected for correlation between the responses of a given subject (subject-level cluster correction).

	Coefficient	
Constant	-1.340	(0.036)
EV/100	1.836	(0.000)
EV/*BO*	-1.188	(0.004)
Stdev/*EV*	2.186	(0.000)
Limelight	0.509	(0.004)
LL	-131.1	
McFadden R^2	0.355	
No. obs.	579	

- *EV*/*BO*: included to control for the expected return of continuing play, and calculated as the average remaining prize divided by the bank offer, or the expected relative return (+1) from rejecting both the current and all subsequent bank offers;

- *Stdev*/*EV*: included to control for the riskiness of continuing play, and calculated by dividing the standard deviation of the distribution of the average remaining prize in the next round by the current average remaining prize;

- *Limelight*: the main variable of interest, a dummy variable that takes the value of 1 if the choice was made in the limelight treatment and 0 if it was made in the lab treatment.

We do not include the common demographic variables *Age* and *Gender*. Our subjects are all students of about the same age, and *Gender* does not have significant explanatory power. We allow for the possibility that errors of individual subjects are correlated through cluster corrections on the standard errors (Wooldridge, 2003).

Table 4.1 shows the regression results. As expected for risk-averse individuals, the propensity to "Deal" is positively related to the riskiness of continuing play, and negatively related to the expected return of continuing play. Furthermore, the "Deal" propensity increases with the stakes. Consistent with the simple analysis of stop rounds, subjects in the limelight are more likely to "Deal" than those in the laboratory (p = 0.004).

As mentioned earlier, in the context of DOND, people have been shown to take more risk after earlier expectations have been shattered or surpassed. In order to investigate this pattern descriptively for our two treatments, we classify subjects as being "losers", "neutrals", or "winners". We follow the method of Post *et al.* (2008), which takes into account the downside risk and the upside potential of rejecting a bank offer. In particular, we define a subject's best-case scenario (BC_r) and worst-case scenario (WC_r) of opening another case in round r as:

$$BC_r = \frac{n_r \bar{x}_r - x_r^{\min}}{n_r - 1} \tag{4.1}$$

$$WC_r = \frac{n_r \bar{x}_r - x_r^{\max}}{n_r - 1} \tag{4.2}$$

where n_r is the number of remaining cases in round r, $\bar{x}_r$ is the average remaining prize in round r, and $x_r^{\min}$ and $x_r^{\max}$ stand for the smallest and largest remaining prize, respectively. A subject is classified as a "loser" if her BC_r belongs to the worst one-third of all subjects in that round, and as a "winner" if her WC_r belongs to the best one-third. Game situations that satisfy neither condition (or both) are classified as "neutral". If two subjects share the same BC_r or WC_r but one falls below the one-third cutoff and one above it, then both are classified as "neutral".

Table 4.2: Decisions after Bad and Good Fortune (First Experiment)

The table summarizes the decisions of the subjects in the laboratory (Panel A; *N* = 40) and limelight (Panel B; *N* = 40) treatment of our first experiment. The samples are split based on the fortune experienced during the game. A subject is classified as a “loser” (“winner”) if her average remaining prize, after eliminating the lowest (highest) remaining prize, is among the worst (best) one-third for all subjects in the same game round. The table displays the percentage bank offer (“%BO”), the number of subjects (“No.”) and the percentage of subjects choosing “Deal” (“%D”) for each category and game round.

	Loser			Neutral			Winner		
Round	%BO	No.	%D	%BO	No.	%D	%BO	No.	%D
A. Laboratory									
1	6	14	0	6	12	0	6	14	0
2	15	14	0	13	12	0	15	14	0
3	42	13	0	32	14	0	32	13	0
4	68	14	0	61	12	8	56	14	7
5	83	13	0	74	12	0	73	13	15
6	92	12	8	88	12	25	86	12	25
7	98	9	33	99	11	0	94	9	11
8	104	6	17	101	13	23	102	6	17
9	101	5	20	102	10	70	104	5	60
1-9		100	6		108	13		100	11
B. Limelight									
1	6	14	0	6	12	0	6	14	0
2	15	14	0	13	12	0	15	14	0
3	42	13	0	32	14	0	32	13	0
4	68	14	0	61	12	25	56	14	14
5	81	12	0	76	11	18	74	12	17
6	92	11	9	87	9	44	88	11	55
7	94	4	25	98	12	25	93	4	0
8	106	4	0	101	8	50	102	4	75
9	108	1	0	101	7	29	105	1	100
1-9		87	2		97	19		87	16

Table 4.2 shows the choices of subjects conditional on the classification of their game situation. In both treatments, winners and losers continue play more often than subjects in the neutral group. This difference is especially pronounced for losers.

Structural Models

We now move to the estimation of structural choice models in order to examine how the more risk-averse behavior in the limelight as opposed to the laboratory corresponds to differences in risk preference parameters, and to further investigate the pattern of path dependence. We implement two simple structural models: one in the spirit of expected utility theory (EU), and the other inspired by prospect theory (PT).[35]

Structural choice models allow for a wide range of specifications. For example, there are many ways to specify the utility function, the error term, reference point dynamics, and probability weighting. We follow the methodology used in the earlier DOND-based studies by Post *et al.* (2008) and Baltussen *et al.* (2012), and summarize this approach below. For further methodological details, background and discussion, we refer to these two prior studies.

For our EU specification, we apply a flexible-form expo-power utility function that allows for the combination of increasing relative risk aversion (IRRA) and decreasing absolute risk aversion (DARA):

$$u(x) = \frac{1 - \exp(-\alpha x^{1-\beta})}{\alpha} \tag{4.3}$$

where α and β are the risk aversion coefficients, subject to $\alpha\beta \geq 0$ to exclude (more exotic) utility functions that combine concavity and convexity. The expo-power function reduces to a CRRA (constant relative risk aversion) power function when $\alpha \to 0$, and to a CARA (constant absolute risk aversion) exponential function when $\beta = 0$.[36]

[35] Henceforth, we refer to these models as the EU model and the PT model. We acknowledge that both theories can be implemented through numerous different and sometimes overlapping specifications. The fit for EU could, for example, be improved with an even more flexible utility function that has both concave and convex segments. As explained in the introduction, our study does not aim to point out whether prospect theory or expected utility theory has greater descriptive power.

[36] We do not follow Post *et al*. (2008) in including initial wealth as a free parameter. This simplification is in line with the standard approach in experimental research, and rules out the possibility of erroneously capturing differences in risk aversion between randomized treatments by differences in wealth estimates.

For PT, we use a simple representation that incorporates loss aversion, uses probabilities as decision weights, and has equal curvature for gains and losses. In particular, the value function is defined as:

(4.4) $$v(x \mid RP) = \begin{cases} (x - RP)^{\alpha} & x > RP \\ -\lambda(RP - x)^{\alpha} & x \leq RP \end{cases}$$

where $\lambda > 0$ is the loss-aversion parameter, *RP* is the reference point, and $\alpha > 0$ represents the curvature of the value function.

Recent literature suggests that reference points are expectation-based and dynamically but partially updated (Kőszegi and Rabin, 2006, 2007; Abeler *et al.,* 2011; Baucells, Weber and Welfens, 2011; Ericson and Fuster, 2011). In this spirit, the reference point in round *r*, RP_r, is modeled as a function of the current bank offer, B_r, and the relative increase of the average remaining prize during the game, $d_r = (\bar{x}_r - \bar{x}_0)/\bar{x}_r$:

(4.5) $$RP_r = (\theta_1 + \theta_2 d_r) B_r$$

where θ_1< 1 (θ_1>1) indicates that the reference point generally takes a value below (above) the current bank offer, and where $\theta_2 \leq 0$ allows for (imperfect) updating of the reference point across rounds. θ_2 = 0 reflects perfect updating, while θ_2 < 0 implies that the reference point sticks to initial expectations. To illustrate: when θ_1 = 1 and θ_2 = 0, the reference point equals the current bank offer; when θ_1 = 1 and θ_2 = -1, the reference point corresponds to the amount that would have been on offer if the average prize had been at its starting level; when θ_1 = 0 and θ_2 = 0, the reference point is zero and all outcomes are considered as gains. Combined with loss aversion and a value function that is concave for gains and convex for losses, the reference point model allows for break-even and house-money effects.

Post *et al.* (2008) and Baltussen *et al.* (2012) also include a separate term for changes during the last two rounds. We drop this short-term lag for brevity and convenience. As also found by Baltussen *et al.* (2012), intermediate changes are economically and statistically insignificant for the reference point, and including the term has no material effect on the other parameters for each of our treatments.[37] Moreover, the use of one single stickiness parameter facilitates comparisons between treatments.

We make the simplifying assumption that subjects look ahead only one round, implying that they compare the current bank offer with the distribution of possible bank offers in the next round. As explained by Post *et al.* (2008), assuming a myopic frame rather than multi-stage backward induction is behaviorally plausible and does not materially affect the results. Post *et al.* (2008) also show that the percentage bank offer can be adequately captured by the simple function:

$$b_{r+1} = b_r + (1 - b_r)\rho^{(9-r)} \tag{4.6}$$

where b_r is the percentage bank offer relative to the expected value of the remaining prizes in round r, and ρ measures the speed at which it approaches the expected value ($0 \leq \rho \leq 1$). Post *et al.* (2008) estimate ρ to be 0.832 for the 40 episodes of the Dutch edition of DOND that we used as scenarios in our experiment. In our analysis, we treat this bank offer model as deterministic and known to the subjects.

We apply maximum likelihood techniques to estimate the unknown parameters. The likelihood of each decision is based on the utility difference between the current bank offer and future bank offers. We assume that a decision is more difficult if the standard deviation of the utility values from continuing play is larger, and set the standard deviation of the model error proportional to this measure. To reduce the potential influence of individual observations, we truncate the likelihood of each decision at a minimum of one percent.

[37] The unimportance of recent changes for the reference point in experiments can be explained by the shorter duration of a game. The original model was designed to capture the behavior of contestants in the TV version of the game, where the recording of a game lasts for about an hour and where recent developments are thus more salient. In our experiments, a game lasts no more than ten minutes, increasing the likelihood that subjects simply compare their current situation with that at the start of their game.

Table 4.3: Expected Utility Model Estimates (First Experiment)

The table displays the maximum likelihood estimation results of our EU model for the laboratory (Panel A; $N = 40$) and limelight (Panel B; $N = 40$) treatment of our first experiment. Shown are the risk aversion parameters (α and β) of the utility function, and the noise parameter (σ). The table also shows the log-likelihood (LL), the AIC and BIC statistics, and the number of decisions (No. obs.). The implied certainty coefficient (CC; certainty equivalent as a fraction of the expected value) is shown for 50/50 gambles of €0 or €10^z, $z = 0, 1, 2, 3$. The *p*-values (within parentheses) are corrected for correlation between the responses of a given subject (subject-level cluster correction).

	Laboratory		Limelight	
α	-		0.021	(0.000)
β	-0.861	(0.022)	0.000	(1.000)
σ	0.544	(0.000)	0.332	(0.000)
LL	-85.5		-78.7	
AIC	177.0		163.4	
BIC	188.2		174.2	
No. obs.	308		271	
CC (0/1)	1.378		0.995	
CC (0/10)	1.378		0.948	
CC (0/100)	1.378		0.554	
CC (0/1000)	1.378		0.067	

Table 4.3 gives the results of the EU model for both treatments. For the laboratory treatment, the expo-power function converges to a risk-seeking CRRA power function. In terms of explanatory power, this model outperforms a naive model that assumes risk neutrality ($\chi^2(2) = 24.27$, $p < 0.001$). In contrast, the function reduces to a risk-averse CARA exponential function for the limelight treatment. This model also fits the data better than a risk-neutral model ($\chi^2(2) = 10.29$, $p = 0.006$).

The shapes of the estimated utility functions are thus very different for the two treatments: one is convex and the other concave. Certainty equivalents (CEs) and certainty coefficients (CCs) can help to interpret the degrees of risk aversion implied by the models. The values nicely illustrate the substantial differences between the two treatments. For a lottery with a 50 percent chance of €100 and €0 otherwise, the implied CE under limelight conditions is €27.72. The CC is 27.72 / 50.00, or 55 percent. For the laboratory treatment, the CE (CC) of €68.91 (138%) is well above the expected value (100%).

Table 4.4: Path Dependence (First Experiment)

The table shows the maximum likelihood estimation results of our EU model for subsamples from the laboratory (Panel A) and limelight (Panel B) treatment of our first experiment. For each treatment, the sample is split based on the fortune experienced during the game. A subject is classified as a "loser" ("winner") if her average remaining prize, after eliminating the lowest (highest) remaining prize, is among the worst (best) one-third for all subjects in the same game round. Definitions are as in Table 4.3.

	Loser		Neutral		Winner	
A. Laboratory						
α	-		-		-	
β	-1.459	(0.070)	-0.459	(0.290)	-0.668	(0.101)
σ	0.530	(0.000)	0.431	(0.007)	0.577	(0.000)
LL	-22.0		-32.9		-28.5	
No. obs.	100		108		100	
CC (0/1)	1.509		1.244		1.320	
CC (0/10)	1.509		1.244		1.320	
CC (0/100)	1.509		1.244		1.320	
CC (0/1000)	1.509		1.244		1.320	
B. Limelight						
α	-2.251	(0.095)	0.018	(0.064)	0.027	(0.000)
β	0.000	(1.000)	0.000	(1.000)	0.000	(1.000)
σ	0.271	(0.000)	0.374	(0.000)	0.252	(0.000)
LL	-8.3		-34.5		-20.2	
No. obs.	87		97		87	
CC (0/1)	1.473		0.995		0.993	
CC (0/10)	1.938		0.954		0.932	
CC (0/100)	1.994		0.595		0.464	
CC (0/1000)	>1.999		0.076		0.051	

The EU specification has difficulties to capture the different preferences of losers, neutrals and winners (as defined earlier). This is illustrated in Table 4.4, which reports separate EU-model estimates for the subsamples. In the limelight, the estimated utility function for losers reflects a preference for risk, while neutrals and winners are risk averse. In the laboratory, each subgroup is best described by a model of risk-seeking preferences, but losers are more risk prone than neutrals and winners. The CCs illustrate the differences between the utility functions.

Table 4.5: Prospect Theory Model Estimates (First Experiment)

The table displays the maximum likelihood estimation results of our PT model for the laboratory (Panel A; $N = 40$) and limelight (Panel B; $N = 40$) treatment of our first experiment. Shown are the loss aversion (λ) and curvature (α) parameters of the value function, the two parameters of the reference point model (θ_1 and θ_2), and the noise parameter (σ). The table also shows the log-likelihood (LL), the AIC and BIC statistics, and the number of decisions (No. obs.). The implied certainty coefficient (CC; certainty equivalent as a fraction of the expected value) is shown for 50/50 gambles of €0 or €10^z, for any $z > 0$, assuming that the reference point equals 0%, 100%, or 200% of the expected value. For λ and α, the null hypotheses are that these parameters equal unity, implying no utility curvature and no loss aversion. The other parameters are tested relative to zero. The p-values (within parentheses) are corrected for correlation between the responses of a given subject (subject-level cluster correction).

	Laboratory		Limelight	
α	0.554	(0.000)	0.711	(0.001)
λ	1.505	(0.042)	2.825	(0.000)
θ_1	1.014	(0.000)	1.040	(0.000)
θ_2	-0.045	(0.001)	-0.072	(0.019)
σ	0.334	(0.000)	0.257	(0.000)
LL	-66.8		-63.6	
AIC	143.7		137.2	
BIC	162.3		155.2	
No. obs.	308		271	
CC (0%)	0.572		0.754	
CC (100%)	0.960		0.796	
CC (200%)	1.428		1.246	

Table 4.5 shows the PT estimates. In the laboratory treatment, we find a rather strong utility curvature, with an α of 0.554. The loss aversion coefficient, λ, equals 1.505. Both values differ significantly from unity (α: $p < 0.001$; λ: $p = 0.042$). Furthermore, the reference point sticks to earlier expectations, with $\theta_2 = -0.045$ ($p = 0.001$). In the absence of changed expectations, it takes a value that is close to the current bank offer ($\theta_1 = 1.014$).

In the limelight treatment, utility curvature ($\alpha = 0.711$, $p = 0.001$) and loss aversion ($\lambda = 2.825$, $p < 0.001$) occur as well. Again, the reference point is sticky ($\theta_2 = -0.072$, $p = 0.019$), and, on average, close to the bank offer ($\theta_1 = 1.040$). While the curvature and reference point parameters are not significantly different between the two treatments, loss aversion is stronger in the limelight than in the laboratory (α: $p = 0.142$; λ: $p = 0.003$; θ_1: $p = 0.166$; θ_2: $p = 0.423$).

Finally, note that the PT model – which can capture the path dependence of risk attitudes through a sticky reference point, loss aversion, and reflection of the value function around the reference point – explains subjects' choices significantly better than the EU model. This better fit holds for both the limelight and the laboratory treatment, also when we take into account the larger number of parameters as compared to the EU model (consider the very different AIC and BIC values).

4.3 Second Experiment

Design and Procedure

To investigate the robustness of the results, we conducted a second experiment. Below we list the design differences. In all other respects, the new experiment was the same as the previous one.

First, we used fixed percentage bank offers. Although subjects in the first experiment had been informed about the two most important factors that determine the bank offer (the bank offer strongly depends on the average remaining prize, and the percentage bank offer gradually increases over the rounds), subjects still faced some ambiguity about the precise offers. Therefore, we cannot exclude that the treatment effects are related to ambiguity rather than risk preferences (Ellsberg, 1961; Camerer and Weber, 1992). In the second experiment we therefore used fixed percentage bank offers for each game round. That is, the bank offer was a percentage of the expected value of the prize in the subject's case that depended on the round number only. For round 1 to 9, the percentages were 15, 30, 45, 60, 70, 80, 90, 100, and 100, respectively. Subjects were informed about this precise structure in the instructions.

Second, we added a third treatment. Subjects in the limelight treatment passively gained experience in playing the game by watching the decisions and outcomes of others. In the first experiment, a subject in the limelight, on average, had watched 9.5 others play the game before she was selected to play herself. In contrast, laboratory subjects did not observe any other subject playing prior to their own game. To examine whether differences in such passive experience matter, we also ran a comeback treatment. This treatment consisted of subjects who had been audience members in the limelight treatment, but had not been selected to play the game on stage themselves. These subjects were invited to play the game in the laboratory afterwards.

An additional benefit of this approach is that subjects in the limelight treatment now always had the opportunity to play the game. In the first experiment, this was not the case as those who were not selected went home empty-handed. As a result, a sense of relief or feelings of luck may have influenced the behavior of those selected. Our announcement of the comeback session avoids this possible confound.

Last, we now used completely random scenarios and more formal experimental instructions. Because comparison with the actual game show was not one of the objectives of this new experiment, there was no need to replay scenarios from the original TV show or to use the instructions that had been handed out to TV show contestants.

The subjects were randomly selected first-year economics students at the Erasmus University of Rotterdam. Subjects who had taken part in the first experiment could not participate. In total, we observed 91 subjects in the laboratory treatment, 40 in the limelight treatment, and 51 in the comeback treatment.[38] All subjects in the comeback treatment had previously watched 20 subjects play the game on stage in our limelight treatment.

[38] The laboratory treatment of this second experiment is also used by Baltussen *et al*. (2012) to investigate different types of incentive systems.

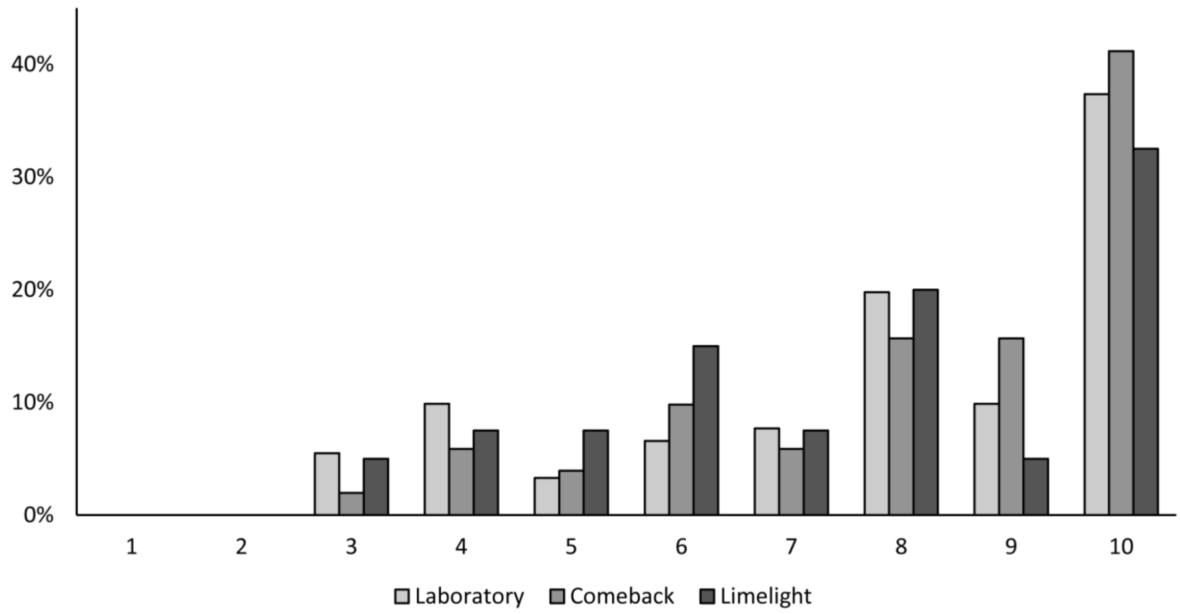


Figure 4.4: Distributions of Stop Rounds (Second Experiment)

The figure depicts the distribution of stop rounds for the three treatments of our second experiment. In the comeback treatment, subjects played the game in a standard economic laboratory setting after viewing others play the game in the limelight treatment. Other definitions are as in Figure 4.3.

Analyses

As with the first experiment, we start with an analysis of the stop rounds.[39] Recall that deciding "Deal" relatively early (late) indicates a relatively high (low) degree of risk aversion. The treatment differences are less pronounced than before. The average stop round in the limelight treatment is 7.55, compared to 7.87 in the laboratory treatment and 8.27 in the comeback treatment. While the average stop round is thus lowest in the limelight treatment, the differences with the two other treatments are not statistically significant (vs. laboratory: *t*-test $p = 0.463$, Mann-Whitney U test $p = 0.406$;

[39] Three subjects in this experiment ended up with trivial choice problems involving prizes of one cent only. Each rejected all nine offers, implying stop round values of 10. The results are not materially different when we set their stop round equal to the number of the first round that had no prizes other than prizes of one cent (or to the average of this number and 10). We omit these uninformative choices in the subsequent probit regression analyses and structural choice model estimations.

vs. comeback: t-test $p = 0.109$, Mann-Whitney U test $p = 0.126$). The difference between the laboratory and the comeback treatment is not significant either (t-test: $p = 0.291$; Mann-Whitney U test: $p = 0.362$). Figure 4.4 shows the distribution of the stop round for the three treatments.

The absence of a treatment effect in the stop rounds may be related to the crudeness of this analysis. While the 40 subjects in each of the two treatments in the first experiment played the same 40 scenarios as TV show contestants, subjects in this experiment faced completely random scenarios. Decision problems can thus be markedly different between treatments, making it even more important to control for differences in the stakes, bank offer, and risk of continuing play. Therefore, we now move to the probit and structural model analyses. For background on the methods, we refer to the previous section.

Table 4.6 shows the results of the probit regression. The results closely resemble those of the first experiment. The propensity to “Deal” is positively related to the riskiness of continuing play and to the stakes, and negatively to the expected return of continuing play. After controlling for these variables, subjects in the limelight turn out to be significantly more likely to “Deal” than those in the laboratory ($p = 0.037$) and then those in the comeback treatment ($p = 0.004$). There is no significant difference between the laboratory and the comeback treatment ($p = 0.414$).

Panel A of Table 4.7 presents the results of the structural model estimations for EU. As in the previous experiment, the expo-power function converges to a risk-seeking CRRA power function for the laboratory treatment. The same is found for the comeback treatment. In both cases the estimated model outperforms a naive model that assumes risk neutrality (laboratory: $\chi^2(2) = 30.23$, $p < 0.001$; comeback: $\chi^2(2) = 15.20$, $p < 0.001$). The β parameter is not significantly different between the laboratory and comeback treatment ($p = 0.359$). For the limelight treatment, the expo-power function again reduces to a risk-averse CARA exponential function that outperforms risk neutrality ($\chi^2(2) = 11.97$, $p = 0.003$).

Table 4.6: Probit Regression Results (Second Experiment)

The table displays the maximum likelihood estimation results of a probit model aimed at explaining the decisions of the subjects in the laboratory (N = 91), limelight (N = 40) and comeback (N = 51) treatment of our second experiment. *Comeback* is a dummy variable that takes a value of 1 for observations from the comeback treatment. Other definitions are as in Table 4.1.

	Coefficient	
Constant	-1.519	(0.000)
EV/100	1.090	(0.000)
EV/*BO*	-0.661	(0.000)
Stdev/*EV*	1.267	(0.000)
Limelight	0.293	(0.037)
Comeback	-0.111	(0.414)
LL	-291.1	
McFadden R^2	0.258	
No. obs.	1367	

Certainty equivalents and certainty coefficients can again help to interpret the parameter values and treatment differences. The CE (CC) for a lottery with a 50 percent chance of €100, for example, is €38.07 (76%) in the limelight. For the laboratory and comeback treatment, the values are €63.07 (126%) and €59.43 (119%), respectively.

Panel B of Table 4.7 presents the estimation results for PT. In the laboratory treatment, we find a rather strong utility curvature, with an α of 0.408. Loss aversion is limited, with λ equaling 1.259. Both values differ significantly from unity (α: $p < 0.001$; λ: $p = 0.005$). Furthermore, the reference point sticks to earlier expectations, with θ_2 = -0.009 ($p < 0.001$), and is, on average, located in the vicinity of the bank offer (θ_1 = 1.002). For subjects in the comeback treatment, we find a curvature of 0.639, a loss aversion of 1.407, and reference point parameters of -0.067 and 1.015 (all $p \leq 0.005$). When we compare the various parameters of these two treatments, we find that subjects in the comeback treatment demonstrate less curvature and a stickier and more elevated reference point than subjects in the laboratory treatment (α: $p = 0.001$; λ: $p = 0.278$; θ_1: $p = 0.014$; θ_2: $p = 0.015$).

Table 4.7: EU and PT Model Estimates (Second Experiment)

The table displays the maximum likelihood estimation results of the EU (Panel A) and PT (Panel B) model for the laboratory (*N* = 91), limelight (*N* = 40) and comeback (*N* = 51) treatment of our second experiment. The panels follow the format and definitions of Table 4.3 and Table 4.5.

	Laboratory		Comeback		Limelight	
A. Expected Utility Theory						
α	-		-		0.010	(0.000)
β	-0.504	(0.000)	-0.332	(0.006)	0.000	(1.000)
σ	0.375	(0.000)	0.296	(0.000)	0.196	(0.000)
LL	-165.2		-88.1		-71.3	
AIC	336.5		182.2		148.6	
BIC	350.0		194.1		159.6	
No. obs.	677		401		289	
CC (0/1)	1.261		1.189		0.998	
CC (0/10)	1.261		1.189		0.975	
CC (0/100)	1.261		1.189		0.761	
CC (0/1000)	1.261		1.189		0.140	
B. Prospect Theory						
α	0.408	(0.000)	0.639	(0.000)	0.751	(0.000)
λ	1.259	(0.005)	1.407	(0.000)	1.863	(0.000)
θ_1	1.002	(0.000)	1.015	(0.000)	1.088	(0.000)
θ_2	-0.009	(0.000)	-0.067	(0.005)	-0.154	(0.000)
σ	0.223	(0.000)	0.163	(0.000)	0.231	(0.000)
LL	-142.7		-59.9		-61.0	
AIC	295.4		129.8		131.9	
BIC	317.9		149.8		150.3	
No. obs.	677		401		289	
CC (0%)	0.367		0.676		0.794	
CC (100%)	0.996		0.951		0.857	
CC (200%)	1.633		1.324		1.206	

In the limelight treatment, we similarly find significant values for utility curvature (α = 0.751, $p < 0.001$), loss aversion (λ = 1.863, $p < 0.001$), and stickiness of the reference point (θ_2 = -0.154, $p < 0.001$). In the absence of changed expectations, the reference point takes a value that is relatively close to the current bank offer (θ_1 = 1.088). In line with the first experiment, subjects in the limelight are more loss averse (p = 0.006) than subjects in the laboratory. In addition they now also display significantly less curvature of the value function ($p < 0.001$) and a stickier and more elevated reference point (both $p < 0.001$). Compared to the comeback treatment, subjects in the limelight are again more loss averse ($p = 0.040$), and they also have a stickier ($p = 0.003$) and more elevated ($p < 0.001$) reference point. Utility curvature is not significantly different ($p = 0.173$).

Similar to the first experiment, the PT model explains subjects' choices significantly better than the EU model. This better fit holds for all three treatments. Further analyses yield evidence of the same pattern of path dependence as in the first experiment: in all treatments, we find that losers have a greater risk appetite than winners and neutral subjects.

The major difference between the two experiments was that the future bank offers were somewhat ambiguous to subjects in the first, and fixed and known to them in the second. A comparison of the estimation results can thus give an indication of whether the effect of ambiguity on behavior is similar or different in and out of the limelight.[40]

Because the bank offer structure differs between the experiments, and the stop round and probit analyses cannot take this difference into account, we consider the structural model results only. For EU, the expo-power function reduces to a similar CRRA power function in the two laboratory treatments; the relevant risk aversion parameter is not significantly different (risk: $\beta = -0.504$; ambiguity: $\beta = -0.861$; $p = 0.371$). In the limelight, however, the risk aversion parameter of the resulting CARA function is marginally significantly larger in the experiment with ambiguity than in the one without (risk: $\alpha = 0.010$; ambiguity: $\alpha = 0.021$; $p = 0.066$). The differences between the CEs (CCs) for a lottery with a 50 percent chance of €100 again illustrate the treatment effects. Under laboratory conditions, the values of €63.07 (126%; risk) and €68.91 (138%; ambiguity) are relatively similar. Under limelight conditions, the values of €38.07 (76%; risk) and €27.72 (55%; ambiguity) are clearly more different.

For PT, the differences in behavior translate into different loss aversion coefficients. For the laboratory treatments, there is no significant difference between the two

[40] Admittedly, comparisons between the limelight treatments are potentially confounded by another design difference. In contrast to subjects in the other treatments, subjects in the limelight treatment with ambiguity (first experiment) were told at the start that only half of them would play the game. A sense of relief or feelings of luck might have influenced the behavior of those selected.

experiments (risk: $\lambda = 1.259$; ambiguity: $\lambda = 1.505$; $p = 0.349$). For the limelight treatments, however, the coefficient is larger in the one with ambiguity (risk: $\lambda = 1.863$, ambiguity: $\lambda = 2.825$, $p = 0.020$).[41]

4.4 Conclusions and Discussion

To analyze how risky choice in the limelight differs from that under more usual experimental laboratory conditions, we conducted two incentivized experiments that mimicked the game of the TV show *Deal or No Deal*. In the laboratory treatments of the experiments, subjects made decisions in a standard, computerized laboratory setting as typically employed in economic experiments. In the limelight treatments, subjects made their choices in a simulated game show environment, which included a live audience, a game show host, and video cameras. The second experiment also had a comeback treatment, in which subjects who had previously gained passive experience by watching others play the game made decisions under laboratory conditions.

We find that subjects are more risk averse in the limelight than in the anonymity of a typical behavioral laboratory. In both experiments, subjects in the limelight have a higher propensity to opt for the sure alternative. For the EU model, this translates into a more concave (risk-averse) utility function. For PT, we observe a higher loss aversion coefficient.

Findings from studies on investor behavior corroborate this result. Barber and Odean (2001, 2002) find that investors trade more and more speculatively after switching from phone-based to online trading. Konana and Balasubramanian (2005) report that investors tend to keep their core investments with traditional brokers and use a small fraction of their wealth to speculate online.

[41] In addition, in the laboratory, the reference point is stickier and more elevated under ambiguity than under risk, and there is marginally significantly less curvature (θ_1: $p = 0.004$; θ_2: $p = 0.009$; α: $p = 0.061$). In contrast, in the limelight, the reference point is less sticky and less elevated under ambiguity than under risk, and there is no significant difference in curvature (θ_1: $p = 0.015$; θ_2: $p = 0.022$; α: $p = 0.707$).

A number of psychological studies suggest that our treatment effect could be related to the emotions evoked by being in the limelight. Emotional states are likely to be different in and out of the limelight, and these differences might bring about differences in risk behavior (Loewenstein *et al.,* 2001; Rick and Loewenstein, 2008). Public scrutiny may entail feelings of stress and anxiety, and lead to a state of physiological arousal. Several studies indicate that anxiety lowers subjects' propensity to take risk. Using both student and clinical samples, Maner *et al.* (2007) show that behavioral and self-report measures of dispositional anxiety are consistently positively associated with risk aversion. Raghunathan and Pham (1999) triggered anxiety in subjects by having them read hypothetical scenarios and asking them to experience these events as vividly as possible by imagining what they would feel and think in these situations. They find that these subjects are more likely to choose a low-risk, low-reward option over a high-risk, high-reward option than subjects in a control group. Kuhnen and Knutson (2011) report similar results. Mano (1994) studies the effect of what he calls "distress" – a combination of a negative (unpleasant) emotional state with a high level of arousal – and finds that subjects who are more distressed display a lower willingness to take risk. Future research could more directly investigate the link between risk tolerance and emotions in and out of the limelight through psychological and physiological measurements of emotional states.

Our second experiment indicates that people in the limelight also have a higher reference point and adjust it more slowly, and that their value function has less curvature. The latter is in line with earlier findings by Miller and Fagley (1991), Takemura (1993, 1994) and Vieider (2011). However, the difference is not significant when we compare subjects in the limelight with experienced subjects in the comeback treatment, suggesting that it may be a spurious effect related to subjects' experience with the game from watching others play. The other results for the comeback treatment reinforce our previous findings about the difference between risk attitudes in and out of the limelight.

While the general degree of risk aversion is affected by the limelight manipulation, the dynamic pattern in risk behavior is not. In particular and in line with the break-even effect, subjects in and subjects out of the limelight are more risk prone when the game develops substantially worse than expected. Of course, on average, losers faced lower stakes, and increasing relative risk aversion (IRRA) thus appears to be a simple explanation for their greater risk appetite. However, IRRA cannot explain that the choice patterns resemble those found by Post *et al.* (2008) for games that used stakes of up to 10,000 times the size of those used in our experiment. Furthermore, IRRA would also imply more risk aversion for winners than for subjects in the middle group, which is not what we observe. The risk appetite of winners is in line with the house-money effect: when all possible outcomes are in the gain domain, people no longer feel they might be losing their “own” money and take more risk.

Our simple PT model allows for a sticky reference point and can capture these path-dependent and very different risk attitudes. All five treatments in our experiments point out that the reference point is sticky and partly determined by subjects’ (presumed) initial beliefs about the task outcome. This finding is in line with recent literature on reference-point formation that argues that reference points are expectation-based and imperfectly updated (Kőszegi and Rabin, 2006, 2007; Abeler *et al.,* 2011; Baucells, Weber and Welfens, 2011; Ericson and Fuster, 2011).

For all treatments, the PT model indeed explains subjects’ choices significantly better than the EU model that we employ. The different degree but similar pattern of risk aversion under the two conditions is important in the light of the recent debate on the external validity of laboratory and field studies (Levitt and List, 2007a, 2007b; Camerer, 2011). Kessler and Vesterlund (2012) argue that while attention has focused on the generalizability of quantitative results, it is much more relevant to focus on the generalizability of qualitative results, as most experimental studies are focused on the direction rather than the magnitude of effects. Furthermore, they argue that while the external validity of quantitative results is highly contested, this is not the case for the external validity of qualitative results. Levitt and List (2007b, p.351) for example state that: “even for those experiments that are affected by our concerns, it is likely that the qualitative findings of the lab are generalizable, even when the quantitative magnitudes are not." Indeed, a large number of studies suggest that qualitative results generalize between lab and field settings, even if quantitative results differ (Kagel and Roth, 2000; Tenorio and Cason, 2002; Healy and Noussair, 2004; Isaac and Schnier, 2005; Antonovics, Arcidiacono and Walsh, 2009; Östling *et al.,* 2011; Bolton, Greiner and Ockenfels, 2013). Our finding of similar patterns of risk taking under different experimental conditions supports the positive view on the generalizability of qualitative results. At the same time, the different degrees of risk taking across conditions sketch a negative picture on the generalizability of quantitative estimates. Where scrutiny has thus far predominantly been considered as a disturbing factor in tasks where moral and wealth are competing objectives (Levitt and List, 2007a, 2007b), this result suggests that scrutiny also affects behavior when moral concerns do not play a role.

The most important difference between the two sets of experiments in our study was that the second used a simple deterministic model for the percentage bank offers that was known to subjects, while subjects in the first set were faced with some uncertainty about the precise offers. Much empirical evidence shows that people are averse to ambiguity, or uncertainty about outcome probabilities (Ellsberg, 1961; Camerer and Weber, 1992). When we compare the results of the two experiments, we find that subjects under limelight conditions are indeed more adventurous when the bank offer structure is deterministic and known rather than ambiguous to them, while we find no evidence that behavior under laboratory conditions is affected by this design change. The different effect of ambiguity in and out of the limelight is in line with literature that suggests that ambiguity aversion is related to the presence of outside observers (Curley, Yates and Abrams, 1986; Trautmann, Vieider and Wakker, 2008; Muthukrishnan, Wathieu and Xu, 2009). Also, the absence of an effect of ambiguity under laboratory conditions corresponds with the findings of Fox and Tversky (1995). Through various experiments under conditions of anonymity resembling our laboratory treatments, they find evidence that ambiguity aversion does not occur when there is no contrast of the ambiguous event with a similar but less ambiguous event. Such a contrast is indeed salient in most studies that classify ambiguity aversion as a real phenomenon. In our case, the task did not embed any contrast, and subjects were not aware of any other related experiment with differently generated bank offers.[42,43]

Comparisons between the results of the comeback treatment and the basic laboratory treatment can identify the effect of passive experience on risk tolerance in our experiment. Recent literature shows that experience helps to eliminate anomalous behavior, in particular loss aversion among market participants (List, 2003, 2004b, 2011; Engelmann and Hollard, 2010; Seru, Shumway and Stoffman, 2010). We find no clear evidence in this direction, perhaps because the experience of the subjects in the comeback treatment was only passive, because learning is slow and subjects observed only 20 others, or because their choice problems were of a different nature than those in a market context. More specifically, we find that the passive experience that comeback-treatment subjects acquired by watching others play does not affect the loss aversion parameter of our PT model, but we do find evidence for decreased curvature and a more elevated and sticky reference point. Interestingly, although the empirical fit of the PT model is much better than that of the EU model for every treatment, the improvement is strongest for the comeback treatment. This suggests that passive experience strengthens rather than weakens prospect-theory like behavior here. A possible explanation is that experience from watching others brings along vivid task-specific expectations and reference, which in turn guide subjects' own behavior. The more sticky reference point of experienced subjects indeed points in this direction.

[42] Interestingly, the comparative ignorance hypothesis of Fox and Tversky (1995) is grounded on the finding of Heath and Tversky (1991) that ambiguity aversion is driven by people's feeling of (in)competence. Possibly, the presence of onlookers in our limelight treatments undermined our subjects' confidence in their capability to perform the task, and this way amplified the effect of the ambiguous bank offer structure on choice.

[43] The uncertainty about the bank offers in our first experiment can be interpreted as a "background risk", although in a strict sense, background risk is mostly regarded and implemented as an additive risk to a subject's overall wealth and not – akin to the uncertainty about future percentage bank offers here – as a multiplicative risk to the outcomes of one choice option only. Based on certain assumptions, most theoretical accounts predict that individuals take less risk in the presence of background risk (Pratt and Zeckhauser, 1987; Gollier and Pratt, 1996; Eeckhoudt, Gollier and Schlesinger, 1996). Experiments by Harrison, List and Towe (2007) and Lee (2008) confirm this prediction, whereas the findings of Lusk and Coble (2008) and Herberich and List (2012) indicate that background risk has little to no effect on risky choice.

Using DOND as the experimental task has a number of advantages, most notably that the game allows for the study of path dependence. However, at the same time, the stop-go nature of DOND might confound the interpretation of our results. In fact, subjects in our limelight treatments were asked to either decide to take risk and stay in the limelight, or to opt for a safe money offer and step out of the limelight. As a result, subjects are more likely to “Deal” if they suffer a fixed disutility from being in the limelight. Although we cannot completely rule out this alternative explanation for our results, it does not appear particularly strong for several reasons. First, self-reflection suggests that such a disutility would rapidly decrease as the game progresses. Many people even get used to being in the limelight after a while. When subjects have to make decisions that make a real difference, they have already gone through an introductory talk with the host and played several trivial game rounds. Second, deciding “No Deal” commits to playing only one round more, and rounds last only briefly, especially at the critical stages of the game when few or only one case is to be opened. The extra time that would be involved is, in fact, negligible in the light of the time already spent on stage. Third, and perhaps most importantly, the data contradict a fixed disutility of being in the limelight. If such a disutility existed, the decisions of the most unfortunate subjects in our sample would be disproportionately strongly affected by it. In our data, we find that losers have a strong tendency to continue play, and this tendency appears to be even stronger in the limelight than in the laboratory, not weaker.

Another potential downside of using DOND is that a game show setting may entail a specific demand effect under limelight conditions. As pointed out by Gertner (1993), taking risk is more entertaining for spectators and this might lead subjects to make riskier choices. In contrast to this intuitive prediction, however, we find that subjects take *less* risk in the limelight than in the laboratory. Apparently, this specific demand effect is relatively unimportant.

In sum, our findings provide a mixed message about the generalizability of findings from one setting to another when the degree of public scrutiny is different. Quantitative measurements of risk preferences do not seem to have universal applicability, but the qualitative pattern of path dependence in risk behavior appears to be robust.

Chapter 5 | On the Social Nature of Eyes: The Effect of Social Cues in Interaction and Individual Choice Tasks

In this chapter, we apply a dual strategy to better understand the effect of pictures of eyes on human behavior. First, we investigate whether the effect of eyes is limited to interaction tasks in which the subjects' decisions influence the outcomes of other subjects. We expand the range of tasks to include individual choice tasks in which the subjects' decisions only influence their own outcomes. Second, we investigate whether pictures of eyes are one of many social cues or are unique in their effect. We compare the effect of pictures of eyes with the effect of a different condition in which we present the subjects with pictures of other students (peers). Our results suggest that the effect of pictures of eyes is limited to interaction tasks and that eyes should be considered distinct from other social cues, such as reminders of peers. While pictures of eyes uniformly enhance pro-social behavior in interaction tasks, this is not the case for reminders of peers. Furthermore, the reminders of peers lead to more rational behavior in individual choice tasks, whereas the effect of pictures of eyes is limited to situations involving interaction. Combined, these findings are in line with the claim that the effect of pictures of eyes on behavior is caused by a social exchange heuristic that works to enhance mutual cooperative behavior.

This chapter is based on the paper "On the social nature of eyes: The effect of social cues in interaction and individual choice tasks.", co-authored by Aurélien Baillon and Asli Selim, and published in Evolution and Human Behavior (Baillon, Selim, and van Dolder, 2013). The authors are grateful to Han Bleichrodt, Rafael Huber, Umut Keskin, Jim Leonhardt, Kirsten Rohde, Joeri Sol, Jan Stoop, Martijn van den Assem and anonymous reviewers for their many constructive and valuable comments on previous versions of this chapter. The chapter benefited from discussion with seminar participants at the Erasmus University of Rotterdam, and with participants of the Tiber Symposium on Psychology and Economics 2011 at Tilburg University, the Subjective Probability, Utility, and Decision Making (SPUDM) 2011 conference in Kingston upon Thames, the Erasmus-Technion Workshop on Decisions and Predictions at Ein Bokek, and the Foundations and Applications of Utility, Risk and Decision Theory (FUR) 2012 conference at Georgia State University.

5.1 Introduction

Humans frequently behave altruistically, even towards genetically unrelated strangers. While some of this altruistic behavior can likely be explained by concerns for the actor's (possible third-party) reputation, it has been argued that this explanation is incomplete. Tightly controlled economic experiments have repeatedly shown that subjects behave in an altruistic manner towards anonymous strangers, even when opportunities for repeated interaction and reputation formation are systematically ruled out (Camerer, 2003). Recent literature, however, has shown that people are sensitive to subtle cues of being watched. In particular, it was demonstrated that, in anonymous experimental settings, the mere presence of pictures of a pair of eyes, or an eye-like stimulus, led to significant increases in donations to strangers in dictator games (Haley and Fessler, 2005; Rigdon *et al.,* 2009; Oda *et al.,* 2011; Nettle *et al.,* 2013), increased donations to a public good (Burnham and Hare, 2007), and induced greater disapproval of moral transgressions (Bourrat, Baumard, and McKay, 2011). The susceptibility of human beings to these subtle cues implies that, even in an anonymous laboratory setting, pro-social behavior should not necessarily be viewed as purely intrinsic (Haley and Fessler, 2005; Jaeggi, Burkhart, and van Schaik, 2010).

A number of studies have investigated the generality of the effect of eyes on social behavior and have attempted to gain deeper insight into the possible mechanisms underlying this effect. A potential concern is that the observed phenomenon may have been caused by an experimenter demand effect (Ekström, 2012). Field experiments, however, suggest that this is not the case, as eye-like stimuli have induced pro-social behavior even when the subjects did not know that they were participating in an experiment. Bateson, Nettle, and Roberts (2006) studied the effect of pictures of eyes on the amount of money that employees at a university psychology department contributed to an “honesty box” in the coffee room. The authors found that, when a picture of eyes was placed next to the “honesty box”, the employee donations tripled. Ernest-Jones, Nettle, and Bateson (2011) showed that placing pictures of eyes in a university cafeteria that required diners to clear their own trays halved the odds of littering. However, the effect of eyes was only significant when the cafeteria was relatively quiet. Similarly, Ekström (2012) found that pictures of eyes increased the amount of money that was donated to charity in Swedish supermarkets by 30% during days on which relatively few people visited the stores. On the days on which the stores were busy, the eyes had no effect on customer donations. Finally, Powell, Roberts, and Nettle (2012) reported similar results to the previous findings. The authors found that displaying pictures of eyes on charity collection buckets in a supermarket increased donations and that this effect was significantly stronger when the supermarket was quiet rather than busy.

Although the eye effect appears to be robust in field settings, several studies suggest that there are conditions under which these effects will not occur. The field studies discussed above suggested that pictures of eyes influence behavior only when the subject is in a non-crowded setting. Fehr and Schneider (2009) found that eyes did not influence the tendency of trustees to repay trust in a trust game. In Mifune, Hashimoto, and Yamagishi (2010), pictures of eyes increased donations in a dictator game when the recipient was an in-group member, but not when the recipient was an out-group member.

The common interpretation of the eye effect is that pictures of eyes trigger feelings of being watched, which in turn activate reputation concerns and subsequent behavioral changes. Such an argument seems plausible, given that actual opportunities to acquire a positive reputation that may pay off in the future have been found to enhance pro-social behavior (Gächter and Fehr, 1999; Wedekind and Milinski, 2000; Milinski *et al.,* 2001; Milinski, Semmann, and Krambeck, 2002; Rege and Telle, 2004; Seinen and Schram, 2006; Engelmann and Fischbacher, 2009). To the best of our knowledge, Oda *et al.* (2011) provided the only direct test of this conjecture. The authors showed that the eye effect was mediated by expectations of future reward but not by a fear of punishment.

In the present chapter, we apply a dual strategy to better understand the effect of eyes on human behavior by expanding both the nature of the tasks and the types of social cues that were used as stimuli. Firstly, we examine whether the influence of eyes is limited to interaction tasks in which the subjects' decisions also influence the outcomes of other subjects, or whether this influence also carries over to individual choice tasks in which the subjects' decisions influence only their own outcomes. There is good reason to believe that eyes may influence decision-making in non-interaction tasks. A long line of psychological research has shown that the mere presence of others can facilitate the performance of simple tasks but impair the performance of more complex tasks (Zajonc, 1965; Bond and Titus, 1983). With respect to choice behavior, research on accountability suggests that people care about how others view their decisions, even in individual choice tasks (Kruglanski and Fruend, 1983; Lerner and Tetlock, 1999; Vieider, 2011). In particular, when subjects know that their decisions will be made public, they adjust their behavior to comply with the prevailing view among their audience. If the view of the audience is unknown, the subjects engage in pre-emptive self-criticism, by carefully analyzing the problem to arrive at a more justifiable decision (Lerner and Tetlock, 1999). These findings are intuitive as people are unlikely to be exclusively concerned with signaling a cooperative disposition; they will, for example, also care about appearing smart, conscientious, and successful. Therefore, if eye-like stimuli trigger a feeling of being monitored, their impact should not be limited to triggering pro-social behavior in interaction tasks, but can be expected to extend to individual choice tasks.

However, it is not definite that the effect of eyes should extend beyond interaction tasks. Cosmides (1989) and Cosmides and Tooby (1989, 1992) argued that humans have evolved specialized, domain-specific cognitive modules for solving problems that are encountered in social exchange. To support this claim, the authors showed empirical evidence that a specialized cheater-detection mechanism existed. Later research suggested that people also have a memory bias for cheaters (see Mealey, Daood, and Krage, 1996, Oda, 1997, and Oda and Nakajima, 2010; see Barclay and Lalumière, 2006, and Mehl and Buchner, 2008, for contradictory findings). The ability to detect and remember cheaters may be necessary to successfully establish relationships of mutual cooperation. However, this ability is not sufficient because people must also aspire to cooperate in the first place. Kiyonari, Tanida, and Yamagishi (2000) therefore proposed the existence of a "social exchange heuristic," which facilitates the establishment of mutual cooperation by encouraging subjects to perceive one-shot prisoner dilemmas as assurance games in which mutual cooperation is the most preferable outcome. As argued by Oda *et al.*

(2011), the eye effect may be due to a similar social heuristic that evolved to facilitate mutual cooperation. If this social heuristic is the cause, then there is no a priori reason to expect pictures of eyes to have any effect in the absence of interaction and thus, no reason to believe that eyes will influence behavior in individual choice tasks.

Secondly, in addition to exploring whether pictures of eyes influenced behavior in individual choice tasks, we investigate the nature of that influence by comparing this effect with the effect of another condition that is designed to remind the subjects of other people in their social group. The literature is somewhat ambivalent regarding whether eyes are special cues or simply one among many social cues that could produce the same result. For instance, in addition to presenting subjects with pictures of eyes, Haley and Fessler (2005) manipulated auditory cues that indicated the presence of others by using sound-deafening earmuffs. The authors found that the earmuffs appeared to reduce the subjects' generosity, although the effect did not reach statistical significance. Lambda and Mace (2010) studied whether the presence of other students influenced decisions in an ultimatum game if the subjects were explicitly guaranteed that their decisions would remain anonymous. The authors found that the presence of other students did not affect the subjects' behavior and cited this result as evidence against an eye effect. Being reminded of others without being exposed to a direct eye gaze may not have the same effect as an eye cue. To investigate whether the effects were the same, we also implement a peers condition in which pictures of our subjects' social group (*i.e.,* university students) are displayed during the experiment.

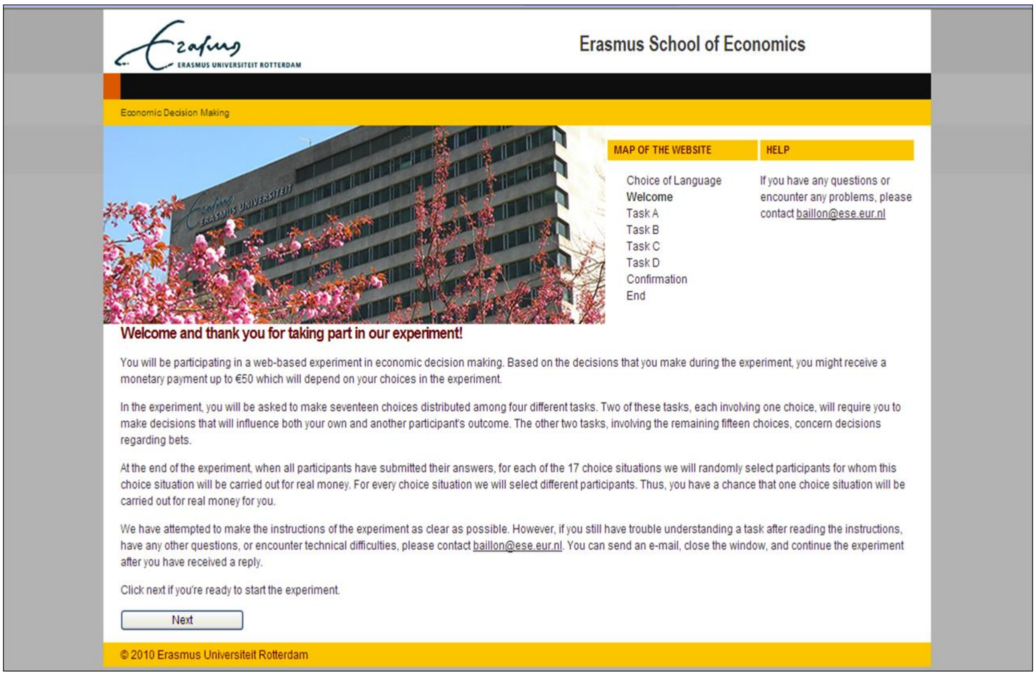


Figure 5.1: Screenshot of the University Website as Used in the Experiment

Placed at the top left of the screen, the pictures randomly rotate every six seconds. The picture displayed on the screenshot above is one of the images that were common to all conditions.

5.2 Method

Subjects

We conducted an online experiment on 165 students from the Erasmus School of Economics (henceforth ESE), Erasmus University Rotterdam, the Netherlands (32% females, age range = 18–33, mean = 21.1 years, S.D. = 2.06 years). The experiment was conducted during the first half of June 2010. We sent an email that contained personalized links to the website developed for the experiment to 600 students. The students were informed that the deadline to participate was two weeks after receipt of the recruitment email and that the payment for their participation could range up to €50; they received an email reminder one week after the initial email. The invitation emails and instructions can be found Appendix 5.A. The subjects were permitted to withdraw from the experiment at any time and their data were analyzed anonymously.

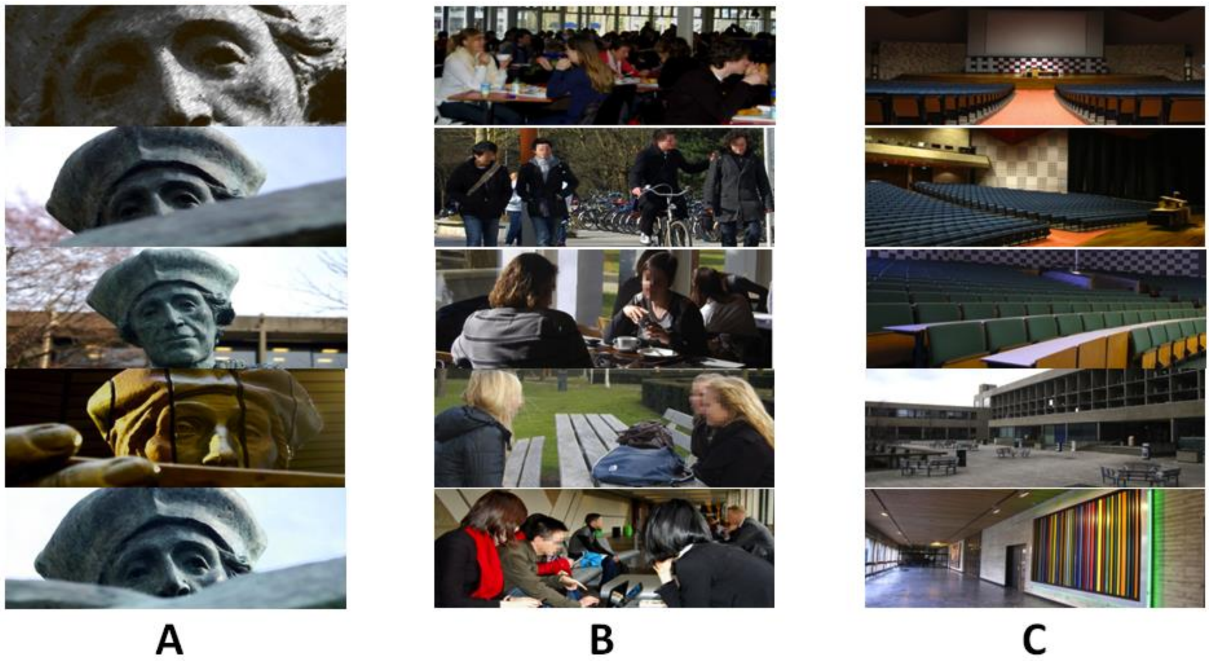


Figure 5.2 Pictures Used in Each Condition.

(A) Eyes, (B) Peers, (C) Control. (In the experiment, the faces of the people in the peers pictures were visible. The faces have been obscured here for publication purposes only.)

Procedure

We constructed a replica of the ESE website (Figure 5.1) for this experiment. After the initial login to any computer at the ESE, Internet Explorer opens up automatically. The homepage consists of the ESE website, which displays news and important information. Students and staff members are required to use this website to look up information and for many administrative procedures. Similarly to the ESE website, our experimental website was bilingual (Dutch and English) and compatible with most browsers (such as Internet Explorer, Mozilla Firefox, Opera, Safari, and Chrome) and most screen sizes.

To present our subjects with pictures of eyes and peers in an unobtrusive manner, we used the picture banner from the official ESE website. This banner typically displays rotating pictures from the campus. The pictures rotate randomly at an approximate interval of six seconds. We constructed three conditions by manipulating the types of pictures that rotated in this banner. The banner was visible to the subjects during the entire experiment.

For the eyes condition, we used pictures of the faces of statues of Erasmus, who is the school's namesake. The students are familiar with images of Erasmus because there are multiple statues of him on the campus and his image appears on official university documents. Thus, using such pictures would not appear out of the ordinary, and we could safely assume that the cues remained sufficiently subtle. Moreover, the neutral facial expressions displayed by the statues reduced the risk of accidently priming emotions (Figure 5.2A).

For the peers condition, we used pictures of students who were not looking directly at the camera to avoid a potential eye effect. The students in these pictures were engaged in studying, chatting, having lunch, etc., on campus. Our subject pool consisted of undergraduate students, thus the representations of their fellow university students could act as social cues that remind them of their own social group (Figure 5.2B, please note that faces have been obscured for publication purposes but were visible in the experiment).

Finally, as a control, we used pictures of empty halls from university buildings (Figure 5.2C). On the whole, the pictures from the three conditions did not differ much from pictures one could find on any university website and were similar to the pictures normally found on the ESE website. In addition to these condition specific pictures, the subjects also viewed two pictures of university buildings that were common to all conditions and were taken from the ESE website. Each subject was randomly allocated to one of the three conditions, and all of the tasks were carried out for real money for some randomly selected subjects after the experiment.

During the experiment, the subjects completed four tasks: two tasks involved interaction between the subjects, and two tasks involved individual choices under uncertainty. The order of the tasks was randomized across subjects. The four tasks were selected on the basis of past research and were designed so that social cues can be expected to impact the subjects' behavior. Each task and the corresponding predictions are described in detail below.

At the end of the experiment, the students answered a small questionnaire including demographic questions (gender, age, nationality, and education). For details we refer to Appendix 5.B. Some of the answers for the first task described below were missing. Approximately 60 subjects were asked to re-enter their answers, of whom 12 failed to do so. As this affected every condition equally, there was no reason to believe that it would affect our results. We nonetheless studied whether it had any effect on our results and found that it had none (see Appendix 5.C.3). For each task, we report simple non-parametric tests for differences between conditions. The more advanced parametric statistical models that control for the subjects' characteristics are reported in Appendix 5.C. All of the results reported in this chapter are robust, and statistical significance is generally stronger in the more advanced analyses than in the simple analyses.

Task 1: Joy of Destruction Mini-Game

The first interaction task we used was the so-called Joy of Destruction mini-game (JoD) (Abbink and Herrmann, 2010). Although research on cooperation and social-preferences has traditionally focused on pro-social behavior, a recent and growing body of literature has begun to apply economic games to the study of anti-social behavior, such as the anti-social punishment of cooperators in public good settings (*e.g.,* Herrmann, Thöni, and Gächter, 2008; Gächter and Herrmann, 2009; Gächter, Herrmann, and Thöni, 2010). The JoD has been used in this literature to show that a considerable fraction of subjects is willing to pay money to destroy part of the payoff to another subject. In particular, the subjects destroyed their opponents' payoffs only infrequently when their behavior could be perfectly observed and their opponents could find out with certainty what caused the destruction. However, when the scenario was altered so that their opponent could no longer find out with certainty whether the destruction was caused by nature or by intention, the subjects' willingness to destroy markedly increased. Note that this difference occurred despite the complete anonymity of the subjects in both cases (Abbink and Sadrieh, 2009; Abbink and Herrmann, 2010).

To achieve a significant amount of destruction and thereby facilitate the investigation of possible differences between our conditions, we adopted the "hidden" setup of the JoD in which it is unclear to the subjects what caused the reduction of their income. In our JoD variant, two subjects each received an endowment of €25. Then, unaware of each other's identity, both subjects were asked whether they would be willing to pay €1 to destroy €10 of the other subject's endowment. There was a 1/3 probability that €10 of the opposing subject's endowment would be destroyed regardless of the subject's decision, making it impossible for the opposing subject to tell what caused the destruction.

In the JoD game, there is no compelling rationale behind destruction: it is harmful to others and costly to oneself. Previous findings on the JoD further suggest that destruction mainly occurs in situations in which the behavior cannot be perfectly observed. In light of these findings, and of past studies that have showed that eyes increase pro-social behavior in simple tasks, we consider this task a way to validate whether the effect of our eyes cues align with the past findings of eyes. Furthermore, the design of this task also allows us to compare the effect of the eyes to the peers condition in an interaction task.

Task 2: Dictator Game

The second interaction task was the dictator game, which is widely studied in economics and which demonstrates what is often deemed to be pure altruism on the part of the subjects (Camerer, 2003). In this game, one subject, the dictator, received a monetary endowment of €50 and was asked how much she would donate to another anonymous subject. The other subject simply received what had been donated to her, and nothing else. The pro-social action here was to donate some money to the receiver, but this would in return lower the dictator's own income. We chose this task because the impact of eye-like stimuli on the dictator game has been studied before (Haley and Fessler, 2005; Rigdon *et al.,* 2009; Oda *et al.,* 2011; Nettle *et al.,* 2013). These past studies found that donation rates were significantly higher in response to eye cues. Including this task in our experiment thus provides us with another opportunity to see whether we could replicate the eye effect in our web-based setup. Furthermore, it provided us with a second opportunity to compare the effect of the eyes to the effect of peers in an interaction task.

Task 3: Ellsberg's Paradox

The third task we employed was a variant of the standard ambiguity aversion task devised by Ellsberg (1961). The task included two bags containing black and red chips. In one bag (Bag K), the proportion of red and black chips was known, whereas in the second bag (Bag U), this proportion was unknown. The subjects were asked to choose a color (black or red) and a bag from which to draw a chip. If the color of the drawn chip matched the color that the subject had chosen, then the subject received €50.

When the proportion of red and black chips is 50-50, Bag K and Bag U are normatively equivalent. Following Laplace's argument that ignorance should be represented by a uniform probability distribution, Bag U should also be considered as a 50-50 bag. If the subjects do not follow this argument and believe that one of the colors makes up more than 50% of the balls in Bag U, then they should bet on this color and strictly prefer Bag U. Nevertheless, many studies have shown that a disproportionate number of people choose Bag K (Camerer and Weber, 1992). The distaste for the unknown bag is often referred to as ambiguity aversion, and, given that the bags are normatively equivalent, can be interpreted as a bias (see, for instance, Raiffa, 1961).

In our experiment, we implemented the standard Ellsberg choice situation with a 50-50 proportion of red and black chips in Bag K, however we also varied the proportion of red and black chips from 10%-90% to 90%-10% (*i.e.,* 10%-90%, 20%-80%, 30%-70%…). For each possible proportion for Bag K, the subjects were asked to state which bag (K or U) they would prefer to draw a ball from. When the probability was different from 50%, the subjects overwhelmingly selected the normatively superior option, *i.e.,* Bag K if the probability of winning in this bag was 60% or higher, and Bag U if the probability of winning in Bag K was 40% or lower. No clear differences between the conditions could therefore be detected in these scenarios (see Appendix 5.C.4). Hence, we report only our analysis of the traditional 50-50 case.

Previous studies have shown that being observed by others matters for this task. Curley, Yates, and Abrams (1986) found that publicly experiencing the consequence of one's own decision in an Ellsberg task generates more ambiguity aversion compared to the situation where privacy was ensured (see also Trautmann, Vieider, and Wakker, 2008, and Muthukrishnan, Wathieu, and Xu, 2009). These authors argued that subjects will fear a negative evaluation if the bet's outcome is not in their favor, and the subjects will believe that choosing bag K is easier to justify due to its informational advantage (its content is known, unlike the one of bag U). Therefore, if our social cues (eyes and peers) trigger concerns of being monitored, we would expect more ambiguity aversion in those conditions compared with the control.

Task 4: Simple vs. Compound Lotteries

Bar-Hillel (1973) has shown that people show systematic biases when comparing simple gambles to compound gambles. To be more specific, people appear to overestimate the likelihood of conjunctive events (*e.g.,* drawing, with replacement, four red chips from a bag with 10 black and 10 red chips) and underestimate the likelihood of disjunctive events (*e.g.,* drawing, with replacement, at least one red chip from a bag with 9 black chips and 1 red chip when the subject is permitted four tries). The cause for this bias is often thought to be a realization of the anchoring and adjustment heuristic (Tversky and Kahneman, 1974). It is believed that, when the subjects evaluate the compound event, they think about the probability of drawing a particular chip, which then takes the role of an anchor. If the subjects do not adjust properly for the compound nature of the event, then they overestimate conjunctive events and underestimate disjunctive events. Thus, people overvalue the conjunctive gambles and undervalue the disjunctive gambles.

In the final task, we investigated the effect of our cues on subjects' evaluation of compound gambles. The subjects were asked to make six choices between simple and conjunctive (compound) gambles. The options presented to the subjects were similar to the ones proposed by Bar-Hillel (1973) and have previously been implemented by Vieider (2011). For instance, in a simple gamble, a subject extracted one chip from a bag that contained 10 red and 10 black chips. The subject received €50 if the chip was red. In the conjunctive, compound gamble, the subject extracted 7 times (with replacement) from a bag that contained 18 red and 2 black chips. The subject won €50 if the chip was red each time. In all of the choice-situations of this task, the probability of winning in the simple gamble exceeded the probability of winning in the conjunctive, compound gamble. Although the simple gamble was thus objectively superior to the compound gamble, past research has showed that a significant number of people found the compound gamble more attractive (Bar-Hillel, 1973; Kruglanski and Fruend, 1983; Vieider, 2011).

In line with the view that lowered anonymity leads to a desire to make better, more justifiable choices, Kruglanski and Freund (1983) and Vieider (2011) found that subjects who expected their choices to be evaluated later on were more likely to make the correct choice when deciding between simple and compound events. Therefore, if our social cues (eyes and peers) triggered the subjects' concerns of being monitored, we would expect them to make fewer mistakes in these conditions compared to the control.

5.3 Results

Task 1: Joy of Destruction Mini-Game

The overall destruction rate obtained in the JoD over the three conditions is similar to the findings in Abbink and Herrmann (2010). Over our entire sample, 24.84 percent of the subjects decide to destroy (N = 153), compared with 25.8 percent of the subjects in Abbink and Herrmann's (2010) experiment. Across conditions, however, we observe sharp differences.

In our control condition (N = 51), the subjects destroy 38.78 percent of the time (Figure 5.3A). The destruction rate is halved in the eyes (N = 49) and peers (N = 53) conditions compared with the control condition, constituting a significant decrease (eyes: 17.65%, $\chi^2(1) = 5.534$, $p = 0.019$; peers: 18.87%, $\chi^2(1) = 4.959$, $p = 0.026$). There is no significant difference between the eyes and the peers condition (χ^2 (1) = 0.026, $p = 0.872$).

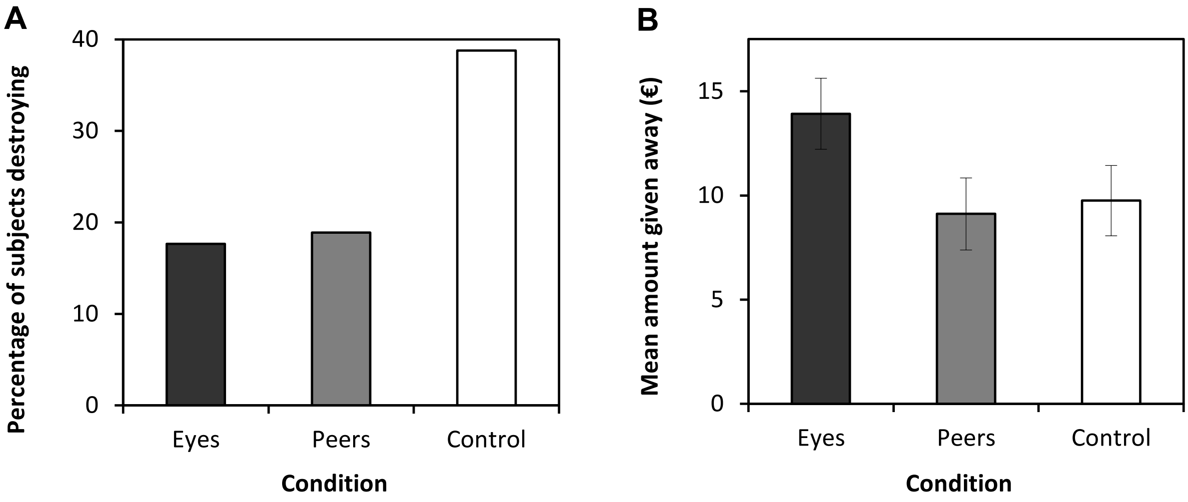


Figure 5.3: Results From the Interaction Tasks

The graph in (A) shows the percentage of subjects who chose to destroy their opposing subject's money in the JoD mini-game, while the graph in (B) shows the mean amount of money that was transferred in the dictator game for the different conditions. Error bars in the graph in (B) show ± 1 standard error of the mean.

Task 2: Dictator Game

The standard finding with respect to the dictator game is that over 60% of the subjects decide to give away money. The mean donation rate across all subjects is typically 20% of the endowment, although the rational, self-interested action is not to allocate any money to the other subject (Camerer, 2003). Across our entire sample, our findings are in line with the statistics presented above; a total of 63.64 percent of our subjects give away money, while the average amount transferred is €10.93, or approximately 22 percent of the €50 endowment (N = 165, 55 in each condition).

In our control condition, the subjects give away €9.75 on average (Figure 5.3B). The pictures of eyes strongly increase donations to an average amount of €13.93 (Mann-Whitney U test p = 0.047). By contrast, the average donation in the peers condition is not significantly different from the control (mean: €9.11, Mann-Whitney U test p = 0.414). The donations amounts are significantly different between the eyes and the peers condition (Mann-Whitney U test p = 0.013).

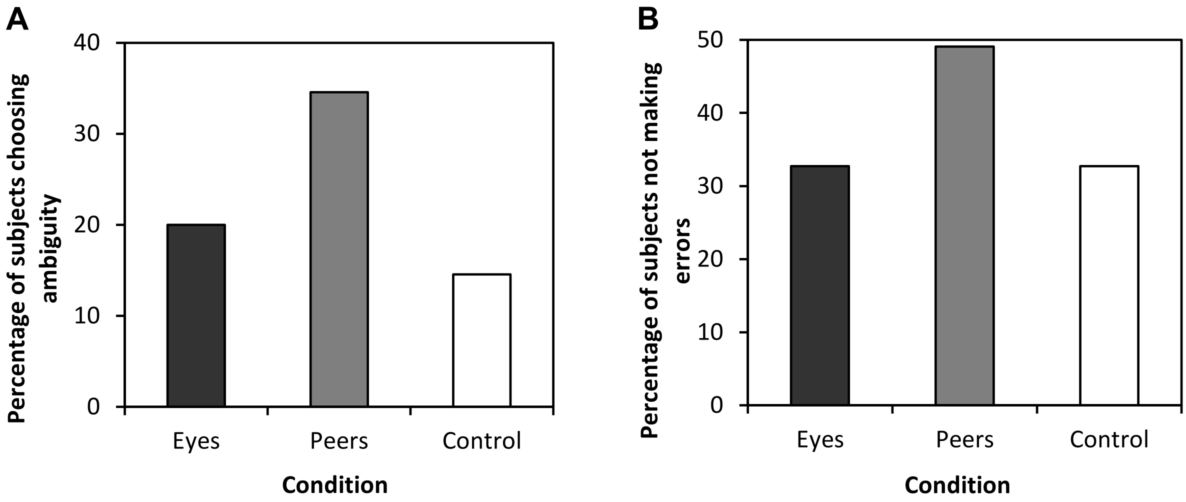


Figure 5.4: Results From the Individual Choice Tasks

The graph in (A) shows the percentage of subjects who chose the ambiguous option (Bag U) over the risky option (Bag K), while the graph in (B) shows the percentage of subjects who did not make any errors in the final task.

Regarding the probability of donating, we find the highest rates of donation in the eyes condition, in which 76.36 percent of the subjects donate some amount. In the control, the percentage of subjects who donate is considerably lower than in the eyes condition, at 63.64 percent, and the lowest rate of donation occurs in the peers condition, at 50.91 percent. Here, however, neither the eyes nor the peers condition differ significantly from the control ($\chi^2(1) < 2.121$, $p > 0.145$). The eyes and peers conditions differ significantly from each other, in that the subjects in the eyes condition are significantly more likely to donate compared the subjects in the peers condition ($\chi^2(1) = 7.700$, $p = 0.006$).

Task 3: Ellsberg's Paradox

In the Ellsberg task, the subjects choose between two bags. The probability of winning was known for Bag K (50%) and unknown for Bag U. In line with past findings, we observe that the majority of subjects choose Bag K in our control condition, while only a small fraction selected Bag U (N = 55, 14.45%, see Figure 5.4A). In contrast to the interaction tasks, we find no effect of eyes on the subjects' bag choice (N = 55, 20%, $\chi^2(1) = 0.573$, $p = 0.449$). In the peers condition, however, the subjects are significantly less likely to show a bias against the ambiguous option than in the other conditions: more than a third of the subjects in this condition choose Bag U (N = 55, 34.55%, comparison with the control: $\chi^2(1) = 5.939$, $p = 0.015$, comparison with the eyes: $\chi^2(1) = 2.933$, $p = 0.087$).

Task 4: Simple vs. Compound Lotteries

The simple gamble is always preferable to the compound gamble; thus, we will refer to the choices that favor the compound gambles as errors. In the control condition (N = 55), fewer than a third of the subjects does not make any errors (Figure 5.4B). There is no difference between the eyes condition (N = 55) and the control condition (both 32.73%, $\chi^2(1) = 0.000$, $p = 1$). In the peers condition (N = 55), however, 49.09 percent of the subjects never make an error. The difference between the peers condition and the two other conditions is marginally significant when the other conditions are separate (both: $\chi^2(1) = 3.046$, $p = 0.081$) and significant at the five percent level when the other two conditions are combined ($\chi^2(1) = 4.160$, $p = 0.041$).

The number of errors reveals a similar pattern to the results presented above. The median number of errors made is one out of six in the peers condition, compared with two out of six in the other two conditions. The mean number of errors made is 2.27 in the control, 1.98 in the eyes condition and 1.60 in the peers condition. Mann-Whitney tests indicate that the difference in the number of errors is marginally significant between the peers and the control conditions (Mann-Whitney *U* test $p = 0.077$). The eyes condition does not differ significantly from the two other conditions ($p > 0.229$).

5.4 Conclusions and Discussion

In the current chapter, we apply a dual strategy to better understand the effect of pictures of eyes on human behavior. First, to identify whether the eye effect is limited to interaction tasks, we expand the range of tasks to include individual choice tasks. Second, to ascertain whether eyes are special or are simply one among many social cues that may produce the same results, we compare the effect of eyes with the effect of another condition that presents the subjects with pictures of other students (peers).

In agreement with past findings, we find that pictures of eyes lead to more pro-social behavior in interaction tasks. Our results reveal that the subjects give more money to strangers and are less likely to destroy the endowment of others in response to eyes cues. However, we find that eyes do not influence subjects' behavior in individual choice tasks, in which their choices do not influence the outcomes of others. This difference suggests that the eye effect is limited to situations that involve interaction, which is compatible with the view that this effect may be caused by a social exchange heuristic that works to establish mutual cooperation, as suggested by Oda *et al.* (2011).

The differences between the eyes condition and the peers condition show that different social cues can have different behavioral implications. In the dictator game, the eyes promote giving, while the peers do not. Moreover, the peers influence behavior in the two individual choice tasks, while the eyes do not. The finding that different social cues can have different effects is important because it implies that care is required to avoid drawing overly general conclusions from the observed effects of one specific social cue.

It is noteworthy that, in the individual choice tasks, the peers condition uniformly increases economic rationality. In that condition, we observe less ambiguity aversion and fewer mistakes in choices between simple versus compound lotteries. In the interaction tasks, we find that peers only influenced behavior in the JoD game, where the pro-social act of not destroying is also economically rational. By contrast, peers do not appear to influence behavior in the dictator game, in which the pro-social and the rational action misalign. In short, the criterion of economic rationality seems to play an important role in the peers condition. It is possible that this effect may be an artifact of our subject pool, which consisted of subjects who were all trained in economics and might fear negative judgment from their peers if they do not make a rational decision. However, it should be noted that this finding also agrees with the general tenet of the accountability literature that considering the judgment of others will encourage pre-emptive self-criticism and careful analysis of the problem to arrive at a more justifiable decision (Lerner and Tetlock, 1999). While the finding of the peers condition in the ambiguity task contradicts the recent literature that suggests that considering others' judgment will increase ambiguity aversion, it is important to note that these papers have all focused on the observation of the actual outcome by others. The accountability literature suggests that expecting judgment based on the outcomes of one’s decisions generally hampers performance, while expecting judgment based on the decision process employed generally improves performance (Simonson and Staw, 1992; Siegel-Jacobs and Yates, 1996). It may be that being presented with pictures of peers during decision-making caused the latter, rather than the former, mechanism to operate. The latter mechanism can explain the results obtained in the present chapter.

It is possible that an alternative mechanism, different from considerations about others’ judgment, may have caused the peers effect. For example, pictures that feature multiple people may trigger a competitive mindset, *i.e.,* a desire to outperform others. Alternatively, the pictures in the peers condition, which displayed other people who did not look directly at the camera, may have made anonymity even more salient than the pictures in the control condition, which did not show any people at all. While the former explanation could account for the increased performance in individual choice tasks, it is not straightforward how the latter could do so. More importantly, both mechanisms fail to account for the findings in the interaction tasks. Competitive subjects should give less than other subjects in the dictator game, which we did not observe. Furthermore, both increased competitiveness and anonymity should be expected to increase destruction in the JoD game. In this game, subjects with a competitive mindset may attempt to improve their relative payoffs by destroying part of their opponents’ endowment, and increasing anonymity has been found to increase destruction rates in previous studies (Abbink and Sadrieh, 2009; Abbink and Herrman, 2010). By contrast, we find that destruction is significantly lower in the peers condition compared with the control.

The influence of our subtle cues on the subjects’ behavior is remarkable, given that the pictures we employed were common pictures that can be found on any university website. Furthermore, it is noteworthy that we find significant effects for both of the social cues in a web-based experiment. Web-based experiments have the advantage of diminishing the participation costs for subjects because they do not need to come to the laboratory and are free to participate at any time. Furthermore, these experiments allow subjects to make decisions in their natural environment. The obvious drawback is that the environment in which subjects make their decisions is less controlled than it would be in the laboratory. For our experiment, it was possible that subjects were in a public setting when they participated in the experiment, which could reduce the relative effectiveness of the social cues (Ernest-Jones, Nettle and Bateson, 2011; Ekström, 2012; Powell, Roberts, and Nettle, 2012). Therefore, using a web-based design instead of a carefully controlled anonymous laboratory setting potentially lowered our chances of finding statistically significant effects (*i.e.,* increased type II errors). That we find statistically significant effects of eyes in both interaction tasks and peers in both the individual choice tasks and one of the interaction tasks suggests that reduction in control was not a major problem in our experiment.

Interestingly, in another recent web-based study, Raihani and Bshary (2012) were unable to find an eye effect in a dictator game played online using Amazon Mechanical Turk (AMT). Our experiment differs from theirs in a number of ways, which makes it difficult to conclusively identify what caused the results to differ. Raihani and Bshary (2012) argued that interacting via AMT may have caused the subjects to feel truly anonymous and therefore be irresponsive to subtle social cues, similar to the argument put forth by Lambda and Mace (2010). This increased anonymity may explain the discrepancy between our findings and the findings from Raihani and Bshary (2012), as AMT ensures a larger degree of anonymity than our experimental setup. In our experiment, the subjects received personalized links to participate in the experiment and the payment of randomly selected subjects was conducted face-to-face so that the subjects could verify that the gambles in individual choice tasks were fairly resolved. Another explanation for the difference, however, may be that in our experiment the subjects played the game against fellow students from the same university, while the subjects in Raihani and Bshary's experiment played against subjects from all over the world. In light of Mifune, Hashimoto, and Yamagishi's (2010) finding that pictures of eyes make people act more altruistically only towards members from their own in-group, this provides another explanation for why we find a significant effect of eyes while Raihani and Bshary (2012) did not.

To study the eye effect in an unobtrusive manner, we used pictures of Erasmus' eyes. Seeing Erasmus on the website would be normal for our subjects, who all studied at the Erasmus School of Economics. However, the image of a famous intellectual such as Erasmus could induce a desire to appear smart. Priming subjects with words such as "professor" has been found to improve subjects' performance at answering trivia questions (Dijksterhuis and van Knippenberg, 1998). Nevertheless, we do not believe that our experiment was compromised in such a way. First, it should be noted that all of the subjects from the three conditions were, in a sense, primed with "Erasmus" because the name Erasmus was displayed at least four times on each screen for each condition (see Figure 5.1, at the top and at the bottom) and on the pictures that were common to all conditions. Moreover, the website that was used closely resembled that of the Erasmus School of Economics. Second, previous research showed that priming subjects with university-related concepts decreased the number of mistakes made by subjects (Dijksterhuis and van Knippenberg, 1998). In our experiment, such priming should mean that subjects should have made fewer errors in the individual choice tasks in the eyes condition. As we have observed, especially in the choices between simple and compound gambles, this reduction in errors did not occur. Pictures of eyes did not lead to better decisions.

Observing that eyes do not influence behavior in our individual choice tasks does not guarantee that eyes will not influence behavior in any individual choice task. It could be argued that subjects react to pictures of eyes only when the task allows them to demonstrate positive qualities, such as being smart, conscientious, or successful, in an obvious manner and that our tasks did not allow them to do so. However, it is important to stress that both of the individual choice tasks were specifically selected to maximize the chance of observing an eye effect. For both tasks, past research indicates that manipulating anonymity in these tasks influences subjects' behavior. Thus, people appear to consider the judgment of others while performing these tasks. Moreover, in the task that compared simple vs. compound lotteries, qualities such as intelligence or conscientiousness could be demonstrated by choosing the objectively superior gamble (all of our subjects had attended mathematical courses on probability theory).

To conclude, our findings suggest that eyes should be considered distinct from other social cues, such as reminders of peers. Although reminders of peers influence a broad range of tasks, the eye effect appears to be limited to triggering pro-social behavior in situations that involve interaction. Combined with findings from previous studies, these results are in line with the claim that responses to eyes are caused by a social exchange heuristic aimed at enhancing cooperative behavior among in-group members (Mifune, Hashimoto, and Yamagishi, 2010), mediated by increased expectations of future reward (Oda *et al.,* 2011).

Appendix 5.A Recruitment Emails

5.A.1 Recruitment e-mail

Dear student,

We would like to invite you to participate in a **web-based** experiment on economic decision-making, run by the "Behavioural Economics Group" at the ESE. The experiment is carried out online, so you can participate at any time and anywhere you like over the *next two weeks*. All you have to do is to use the link below and follow the instructions on the website. The experiment will take 10-15 minutes of your time, and in return you will get a chance to win up to 50 euros! We will randomly select 19 people among the participants and have a budget of 850 euros for this experiment.

Your personal link to the experiment is:

[PERSONALIZED LINK TO THE WEBSITE]

You will not have to log into our website: this personal link will automatically register that you have participated in the experiment.

The experiment will be online only 2 weeks, so if you want to have a chance of winning 50 euros, you should make sure to participate in the experiment very soon.

Thank you for your interest in our experiments!

Best regards,

XXXXXXXXXXX

5.A.2 Reminder e-mail

Dear student,

There is only one week left to take part in our web-based experiment and to get a chance to win 50 euros.

Your personal link to the experiment is:

[PERSONALIZED LINK TO THE WEBSITE]

More information:

This is a **web-based** experiment on economic decision-making, run by the "Behavioural Economics Group" at the ESE. The experiment is carried out online, so you can participate at any time and anywhere you like. All you have to do is to use the link above and follow the instructions on the website. The experiment will take 10-15 minutes of your time, and in return you will get a chance to win up to 50 euros! We will randomly select 19 people among the participants and have a budget of 850 euros for this experiment.

You will not have to log into our website: this personal link will automatically register that you have participated in the experiment.

The experiment will be online only 1 more week, so if you want to have a chance of winning 50 euros, you should make sure to participate in the experiment very soon.

Thank you for your interest in our experiments!

Best regards,

XXXXXXXXXXX

Appendix 5.B Experimental Instructions

5.B.1 Welcome page

Welcome and thank you for taking part in our experiment!

You will be participating in a web-based experiment in economic decision-making. Based on the decisions that you make during the experiment, you might receive a monetary payment up to €50 which will depend on your choices in the experiment.

In the experiment, you will be asked to make seventeen choices distributed among four different tasks. Two of these tasks, each involving one choice, will require you to make decisions that will influence both your own and another participant's outcome. The other two tasks, involving the remaining fifteen choices, concern decisions regarding bets.

At the end of the experiment, when all participants have submitted their answers, for each of the 17 choice situations we will randomly select participants for whom this choice situation will be carried out for real money. For every choice situation we will select different participants. Thus, you have a chance that one choice situation will be carried out for real money for you.

We have attempted to make the instructions of the experiment as clear as possible. However, if you still have trouble understanding a task after reading the instructions, have any other questions, or encounter technical difficulties, please contact XXXXX@ese.eur.nl. You can send an email, close the window, and continue the experiment after you have received a reply.

Click next if you're ready to start the experiment.

Next

5.B.2 Joy of Destruction Mini-Game

In this task, you will be randomly matched with another participant in the experiment. We will refer to this other participant as "Player B". Both you and Player B will receive an endowment of € 25. You have to decide whether to reduce Player B's income or to leave it as it is. If you pay € 1, you can reduce Player B's income by € 10. Player B will be asked to make the same choice regarding your income and will incur the same cost (€ 1) if (s)he chooses to reduce your income.

After Player B and you have decided whether or not to reduce each other's income, a die will be thrown twice. Once for you and another time for Player B.

Let us consider the throw concerning Player B's income. If the die shows 1 or 6 Player B's income will be reduced, independent of your decision. If the die shows any other number (2,3,4,5) then your decision will be realized: If you have decided to reduce Player B's income, the income will be reduced. If you have decided not to reduce Player B's income, the income will not be reduced.

The same procedure will be applied to determine your income: first a throw of a die, then, if the die shows a 1 or a 6, your income will be reduced irrespective of Player B's decision. If the die does not show a 1 or a 6, Player B's decision regarding your income will be carried out.

Please be aware that neither Player B nor you will learn about the outcome of the throws of the die. Therefore, if Player B's income is reduced by € 10, Player B will never learn what the reason for this reduction has been: your decision or the results of the throw of the die. Similarly, if your income is reduced, you will not know whether this is due to Player B's decision or the throw of the die.

Please make your decision: Your endowment in this experiment is € 25.

Do you want to pay € 1 to reduce Player B's income by € 10?

Yes

No

Once you have made your decision, click next.

Next

5.B.3 Dictator Game

You have been allocated € 50. Your task is to decide how much of this amount to allocate to another individual. The other individual will receive this amount and you will keep the rest.

The other individual will be a randomly selected participant of the experiment. This participant cannot be selected to be paid out for his or her own decisions in the experiment; hence, his or her payoff solely depends on your choice. If you happen to be the randomly selected participant whose choice will be paid out for real, we will make sure that you and the other participants will be invited to receive your payments on different days, so as to rule out any chance that you will meet the other participant. You will not learn the identity of the participant you are matched with, and likewise the other participant cannot learn your identity.

You are now asked to state the amount you wish to allocate to the other participant. This must be a number (integer) between 0 and 50.

Once you have made your decision, click next.

Next

5.B.4 Ellsberg Tasks

This task involves 9 choices. For each of these choices, one participant will be randomly selected, and his/her decision will be implemented for real, and the resulting outcome will be paid in euros. Please state your decision for each of the following choice tasks.

You will have to pick a colour: red or black, and draw a chip from a bag containing red and black chips. If your colour is drawn you will win €50, but if the other colour is drawn, you will win nothing. You have to decide from which bag you would like to draw a chip: Bag A or Bag B.

- In Bag A, there will be 10 chips. Each chip can only be black or red, but the proportion of each colour will be unknown. The bag will be ready before you choose your colour, but you will not be allowed to check what is in it before choosing a colour and drawing a chip.

- In Bag B, we will put (in front of you) x chips of your colour and 10 - x chips of the other colour.

If x were 0, Bag A would be more interesting because there could be at least one chip with your colour in this bag. If x were 10, Bag B would be more interesting because it would guarantee €50. For x=1, 2, ..., 9, you have to choose the bag from which you want to extract a chip so as to win €50 if you draw a chip of your colour.

	Choice 1 x=1	Choice 2 x=2	Choice 3 x=3	Choice 4 x=4	Choice 5 x=5	Choice 6 x=6	Choice 7 x=7	Choice 8 x=8	Choice 9 x=9
Bag A	unknown proportions of red and black chips	unknown proportions of red and black chips	unknown proportions of red and black chips	unknown proportions of red and black chips	unknown proportions of red and black chips	unknown proportions of red and black chips	unknown proportions of red and black chips	unknown proportions of red and black chips	unknown proportions of red and black chips
Bag B	1 chip of your colour, 9 chips of the other colour	2 chips of your colour, 8 chips of the other colour	3 chips of your colour, 7 chips of the other colour	4 chips of your colour, 6 chips of the other colour	5 chips of your colour, 5 chips of the other colour	6 chips of your colour, 4 chips of the other colour	7 chips of your colour, 3 chips of the other colour	8 chips of your colour, 2 chips of the other colour	9 chips of your colour, 1 chip of the other colour
	A B	A B	A B	A B	A B	A B	A B	A B	A B

Once you have made your decision, click next.

Next

5.B.5 Compound vs. Simple Lotteries

This task involves 6 choices. For each of these choices, one participant will be randomly selected, and his/her decision will be implemented for real, and the resulting outcome will be paid in euros. Please state your decision for each of the following choice tasks.

Each of the choice tasks involves choosing between an option that involves drawing one chip from a bag and another option that involves drawing multiple chips from a different bag.

In case of drawing multiple chips from the bag, the poker chips you draw will be placed back in the bag and the chips in the bag will be mixed before you extract again, so as to keep the composition of the bag constant. This holds true for all choice situations below.

Please pay attention to both the composition of the bags and the number of extractions, which both vary across tasks.

In each choice situation, you have to choose between two options to win €50.

	Choice 1	Choice 2	Choice 3	Choice 4	Choice 5	Choice 6
Option A	extract 1 time from a bag with 10 red and 10 black chips, win if red	extract 1 time from a bag with 5 red and 15 black chips, win if red	extract 1 time from a bag with 5 red and 15 black chips, win if red	extract 1 time from a bag with 2 red and 18 black chips, win if red	extract 1 time from a bag with 4 red and 16 black chips, win if red	extract 1 time from a bag with 6 red and 14 black chips, win if red
Option B	extract 7 times from a bag with 18 red and 2 black chips, win if 7 times red	extract 5 times from a bag with 15 red and 5 black chips, win if 5 times red	extract 7 times from a bag with 16 red and 4 black chips, win if 7 times red	extract 4 times from a bag with 10 red and 10 black chips, win if 4 times red	extract 6 times from a bag with 15 red and 5 black chips, win if 6 times red	extract 2 times from a bag with 10 red and 10 black chips, win if 2 times red
	A B	A B	A B	A B	A B	A B

Once you have made your decision, click next.

Next

5.B.6 Confirmation Screen and Additional Questions

Confirmation

Your choices have been registered. Please answer the following questions to validate your participation in the experiment.

Read each of the following statements carefully and indicate how characteristic it is of you according to the following scale[44]:

	Not at all characteristic of me	Slightly characteristic of me	Moderately characteristic of me	Very characteristic of me	Extremely characteristic of me
I worry about what other people will think of me even when I know it doesn't make any difference.					
I am unconcerned even if I know people are forming an unfavorable impression of me.					
I am frequently afraid of other people noticing my shortcomings.					
I rarely worry about what kind of impression I am making on someone.					
I am afraid that others will not approve of me.					
I am afraid that people will find fault with me.					
Other people's opinions of me do not bother me.					
When I am talking to someone, I worry about what they may be thinking about me.					
I am usually worried about what kind of impression I make.					

[44] This questionnaire was intended to measure fear of negative evaluation (Leary, 1983). We have decided to disregard this questionnaire for two reasons. First, subjects complained about it in their comments after the experiment (whereas most of the other comments were positive). The main problems seemed to be that these were the only psychological questions we used, which made it overtly obvious to the subjects what we were trying to measure and that the Dutch version of the scale was completely unidirectional (none of the questions were reversely coded, a higher score always implied more fear). Therefore, subjects considered the questions to be suggestive and disliked providing answers. Second, and possibly related to the first point, we noticed in the website's log files that many subjects preferred not to answer these questions and only did so when they were asked to do it for a second time. Therefore, it is very likely that they did not put much effort into answering the questions.

If I know someone is judging me, it has little effect on me.					
Sometimes I think I am too concerned with what other people think of me.					
I often worry that I will say or do the wrong things.					

Did you use a calculator to make some choices in the experiment?

Yes

No

Please indicate you age, gender, year of study, and nationality.

Age:					
Gender:	Male	Female			
Year of study:	Bachelor 1	Bachelor 2	Bachelor 3	Master	Other
Nationality:	Dutch	Other			
Any comment? (optional)					

5.B.7 Final Screen

Thank you for your participation. Your answers have been recorded.

When the experiment is over, we will let you know whether you have been selected to play one of your choices for real.

Appendix 5.C Additional Analyses

5.C.1 Descriptive Statistics

Table 5.1 displays the descriptive statistics for the 162 subjects that completed the questionnaire at the end of the experiment (three subjects neglected to do so). The majority of our subjects were male (68%), Dutch (65%), and in the second year of their bachelor's degree (57%). Furthermore, around 14 percent were in their first year of the bachelor's degree, 12 percent were in their third year, 16 percent were following a master's program, and 1 percent did not fall into any of these categories. Both the average and the median age were 21, and age ranged from 18 to 33. It should, however, be mentioned that over 90 percent of our subjects was under the age of 25 (not in table). A considerable share of subjects admitted to having used a calculator during the experiment (45%). Note that using a calculator was by no means forbidden in the experiment. We simply asked this question since using a calculator would facilitate finding correct answers in one of the tasks.

5.C.2 Joy of Destruction Mini-Game

In the JoD mini-game, the subjects had the option to pay €1 to destroy €10 of another player's endowment. As presented in the chapter, χ^2-tests show that subjects are significantly less likely to destroy the endowment of the other subject in eyes condition and the peers condition, relative to the control. There is no difference between the eyes and the peers condition. In the current section we show that these results are robust when we apply Probit regressions and control for the effect of other variables. Furthermore, due to a technical problem, the decisions submitted by some of the subjects (58 out 165) were initially not recorded in the database. These subjects received an email telling them that they could go back to the website to fill in the missing decision and most of them (46 out of 58) did so. Here, we provide additional tests showing that there is no indication that this data problem influenced our results.

Table 5.1: Descriptive Statistics of the Subjects

The table shows descriptive statistics for our sample of 162 subjects who participated in the experiment and answered the questionnaire (three subjects neglected to do so). *Age* is the subject's age measured in years. *Gender, Nationality, Calculator*, *Bachelor 1*, *Bachelor 2*, *Bachelor 3*, *Master*, and *Other* are dummy variables indicating whether a contestant is female (*Gender*), Dutch (*Nationality*), indicates having used a calculator (*Calculator*), is a first year Bachelor student (*Bachelor 1*), a second year Bachelor student (*Bachelor 2*), a third year Bachelor student (*Bachelor 3*), a master student (*Master*), or indicates that she is neither in the first three years of her Bachelor, nor a Master student (*Other*), respectively.

	Mean	Median	Stdev	Min	Max
Age	21.10	21	2.06	18	33
Gender (Female = 1)	0.32	0	0.47	0	1
Nationality (Dutch = 1)	0.65	1	0.48	0	1
Calculator (yes = 1)	0.45	0	0.50	0	1
Year of study					
Bachelor 1	0.14	0	0.35	0	1
Bachelor 2	0.57	1	0.50	0	1
Bachelor 3	0.12	0	0.32	0	1
Master	0.16	0	0.37	0	1
Other	0.01	0	0.11	0	1

Table 5.2 shows the results for Probit modesl on the probability that a subject destroys the endowment of another subject; significance levels are based on robust standard errors. Furthermore, since coefficients in a Probit model do not offer intuitive interpretations in terms of effect size, we report marginal effects evaluated at the covariate means. Model 5.1 provides a simple comparison between conditions. In line with the χ^2-tests, we observe that destruction rates both in the eyes (p = 0.01) and the peers (p = 0.02) conditions are significantly lower than in the control condition. There is no significant difference between the eyes and the peers condition (p = 0.87, untabulated). Adding our control variables (Table 5.2, Model 5.2), we find that only nationality has a significant influence on destruction. Dutch students are significantly less likely to destroy the other's endowment (p < 0.01). The effect of both the eyes and the peers condition remain statistically significant (respectively p = 0.01 and p = 0.04). In short, these analyses show that the simple, non-parametric tests applied in the main chapter prove robust in more advanced analyses controlling for age, gender, nationality, education year, and the use of a calculator.

Table 5.2: Probit Regression Results on Destruction Rate in the JoD Mini-Game

The table displays results from the Probit regression analyses of subjects' decisions to destroy (1) or not destroy (0) part of another subject's endowment. *Eyes* and *peers* are dummy variables taking the value 1 if subjects were in the eyes condition (*Eyes)* or the peers condition (*Peers*), respectively. Definitions of the other variables are as in Table 5.2. Model 5.1 is estimated on the entire sample of 153 subjects who successfully submitted a decision in the JoD mini-game. Model 5.2 and Model 5.3 are estimated on the set of 150 subjects who both successfully submitted a decision in the JoD mini-game and answered the questionnaire at the end of the experiment. Model 5.4 is estimated on the set of 104 subjects for whom the decision in the JoD mini-game was successfully recorded the first time round (i.e. not affected by technical problems) and who answered the questionnaire at the end of the experiment. For each explanatory variable, the marginal effect at the covariate means is shown. Robust standard errors are used and *p*-values are shown in parentheses.

	Model 5.1		Model 5.2		Model 5.3		Model 5.4	
Condition dummies								
Eyes	-0.18	(0.01)	-0.18	(0.01)	-0.18	(0.01)	-0.18	(0.03)
Peers	-0.17	(0.02)	-0.14	(0.04)	-0.15	(0.03)	-0.15	(0.08)
Control variables								
Age			-0.01	(0.50)	-0.01	(0.55)	0.00	(0.90)
Gender (female = 1)			0.06	(0.46)	0.07	(0.40)	0.05	(0.64)
Nationality (Dutch = 1)			-0.26	(0.00)	-0.26	(0.00)	-0.30	(0.00)
Year of study								
Bachelor 2			0.02	(0.86)	-0.04	(0.76)	-0.23	(0.24)
Bachelor 3			0.04	(0.80)	0.01	(0.93)	-0.21	(0.00)
Master			0.18	(0.39)	0.14	(0.49)	-0.04	(0.87)
Calculator (yes = 1)			-0.06	(0.43)	-0.06	(0.42)	0.01	(0.93)
Second time (yes = 1)					-0.08	(0.27)		
Log pseudo-likelihood		-82.15		-69.21		-68.74		-47.2
N		153		150		150		104

Table 5.3 shows descriptive statistics on destruction rates depending on whether the subject's decision was recorded the first time round or whether they had to record their decisions a second time due to the data storage problem. For the eyes and the control condition, the findings are highly similar in both cases. Subjects destroy the others' endowment in 18.18 percent of the cases when recording their decision for the first time and 16.67 percent when recording it the second time in the eyes condition. These statistics are 39.39 percent and 37.50 percent, respectively, for the control condition. When investigating the peers condition the gap appears a bit larger, subjects destroy the other's endowment in 21.95 percent of the cases when answering the question for the first time, and 8.33 percent of the cases when answering the question for a second time. A Fisher's exact test, however, indicates that this difference is not statistically significant ($p = 0.42$).

Table 5.3: Descriptive Statistics on Destruction Rate

The table shows descriptive statistics on destruction rate in the JoD mini-game depending on whether subjects' decisions were recorded the first time the subjects submitted them, or whether subjects had to record their decisions for a second time due to a technical problem with the website. Results are shown for all conditions both separately and combined. Overall statistics are provided in the final column.

	First time		Second time		Overall	
	N	% destroy	N	% destroy	N	% destroy
Eyes	33	18.18	18	16.67	51	17.65
Peers	41	21.95	12	8.33	53	18.87
Control	33	39.39	16	37.50	49	38.78
Total	107	26.17	46	21.74	153	24.84

Overall, the qualitative pattern seems to be the same independent of whether subjects recorded their decisions for the first or the second time: subjects in the eyes and peers conditions destroy at a similar rate, which is lower than the destruction rate in the control condition. Performing χ^2-tests, we find that even for the subset of subjects who recorded the questions for the first time the differences between conditions approach significance (comparing eyes with control: p = 0.06; comparing peers with control: p = 0.10). The additional observations from the subjects who had to record their decision for a second time thus only strengthen the statistical evidence for an already apparent pattern.

Table 5.2, Model 5.3 incorporates a dummy variable taking the value 1 if the decision had to be recorded a second time, 0 otherwise, into the full model. This analysis shows that being requested to answer the question a second time does not steer behavior in a particular direction. Model 5.4 reports the results of estimating the model on the subset of subjects whose decisions were successfully stored the first time round. While the significance levels drop a bit, we observe that the estimates of the marginal effects are not at all affected by leaving out these subjects. This provides further indication that the data storage problem did not affect our results in a meaningful way.

5.C.3 Dictator Game

In the main chapter, we have shown that the results from non-parametric Mann-Whitney tests suggest that the amount donated in the dictator game differs significantly between conditions. In particular, subjects donate significantly more to the other subject in the eyes condition compared to the other two conditions. There is no significant difference in the amount donated between the peers and the control condition. Looking at the percentage of subjects who decide to give away money, χ^2-tests indicate that neither the eyes nor the peers conditions differs significantly from the control, but that subjects in the eyes condition are significantly more likely to donate as compared to subjects in the peers condition. In the present section we will show that these results are robust, or even strengthened, by performing more advanced analyses and controlling for the effect of other variables on the willingness to donate money to a stranger.

First, we analyze the amount donated by the Dictator by means of a Tobit model. We use a Tobit model to account for the fact that our dependent variable "Amount given" is censored between €0 and €50. Model 5.1 and Model 5.2 in show our results, significance levels being based on robust standard errors. Model 5.1 presents a simple test of the condition effects to compare the results from the (parametric) Tobit analyses with those of the (non-parametric) Mann-Whitney tests reported above. We observed that the difference between the eyes condition and the control condition decreases in significance due to the distributional assumptions made in the Tobit. Still, the difference between the eyes and the control conditions remains marginally significant (p = 0.08)—and is significant if we perform a one-sided test (p = 0.04). More importantly, however, adding our control variables (Model 5.2) the condition effect increases in significance, becoming significant at the 5 percent level in a two-sided test (p = 0.04). None of the control variables seems to have a strong influence on behavior, except that Dutch students seem less willing to donate money (p = 0.09).

Table 5.4: Regression Results on Giving in the Dictator Game

The table displays results from regressions on the subjects' giving behavior in the dictator game. Model S4.1 and Model 5.2 display results of Tobit regression analyses on the amount donated by the subjects. Model 5.3 and Model 5.4 displays results of Probit analyses on the subjects' decisions to either donate (1) or not (0). Definitions of the variables are as in previous tables. In the results of the Probit regressions we depict marginal effects at the covariate means. For both Tobit and Probit models, we apply robust standard errors. *P*-values are shown in parentheses.

	Amount transferred				Probability giving			
	Model 5.1		Model 5.2		Model 5.3		Model 5.4	
Constant	5.15	(0.06)	-3.16	(0.91)				
Condition dummies								
Eyes	6.31	(0.08)	7.41	(0.04)	0.13	(0.13)	0.18	(0.05)
Peers	-2.91	(0.48)	-2.81	(0.51)	-0.12	(0.18)	-0.09	(0.38)
Control variables								
Age			0.88	(0.50)			0.01	(0.72)
Gender (female = 1)			-3.04	(0.37)			0.04	(0.64)
Nationality (Dutch = 1)			-5.62	(0.09)			-0.22	(0.01)
Year of study								
Bachelor 2			-5.29	(0.29)			-0.10	(0.39)
Bachelor 3			-4.92	(0.39)			0.06	(0.73)
Master			-7.31	(0.35)			-0.17	(0.41)
Calculator (yes = 1)			-3.60	(0.31)			-0.02	(0.85)
Sigma		18.80		18.51				
Log pseudo-likelihood		-490.89		-474.73		-104.24		-97.92
N		165		162		165		162

Interpreting the parameters, an individual's willingness to donate increases by about €7.41 euro's in the eyes condition compared to the control condition. This difference is larger than the observed difference in money allocated between conditions due to the fact that the Tobit takes censoring in the data into account. As could be expected, the difference between the eyes condition and the peers condition is significant both in Model 5.1 and Model 5.2 ($p < 0.02$, untabulated).

Model 5.3 and 5.4 show the results of Probit models on the probability that a subject allocates a non-zero amount to another subject. As before, we report marginal effects evaluated at covariate means and significance levels are based on robust standard errors. Model 5.3 provides a simple comparison between conditions. As suggested by the χ^2-tests reported earlier, the Probit results shows that no condition differs significantly from the control condition, while the subjects in the eyes condition are significantly more likely to donate compared to those in the peers condition ($p < 0.01$, untabulated). Adding our control variables, however, increases the significance of the eyes condition sharply, indicating that subjects in the eyes condition are significantly more likely to give a positive amount to another subject compared to the subjects in the control condition (p = 0.05). The size of this effect is impressive: the subjects in the eyes condition are almost 18 percentage points more likely to donate money compared to the subjects in the control condition and more than 25 percentage points more likely to donate money compared to the subjects in the peers condition (p < 0.01). The difference between the peers condition and the control condition remains insignificant (p = 0.38). Again the only control variable that seems to matter is nationality, Dutch students are 22 percentage points less likely to allocate a positive amount to another subject (p = 0.01).

In conclusion, using more advanced analyses and controlling for a range of other variables that can potentially influence giving behavior, we find that this only strengthens the conclusions drawn in the main chapter. It is interesting to note that while Dutch students acted less anti-social in the JoD mini-game, these same students acted less pro-social in the Dictator game. This suggests that this subcategory of students is not, in fact, more or less kind, but rather is less likely to deviate from the prediction of rational self-interest.

5.C.4 Ellsberg Tasks

Here we present a number of additional analyses regarding the ambiguity questions used in the experiment. First, we will show that the findings in the general Ellsberg tasks (Ellsberg, 1961) that are reported in the main chapter are robust if we use Probit analyses and control for the effect of other variables on the subjects' decisions. Second, we will investigate the other questions posed to subjects. As mentioned in the main chapter, we implemented the standard Ellsberg choice situation with a 50-50 proportion of red and black chips in Bag K, but we also varied the proportion of red and black chips from 10%-90% to 90%-10% (i.e., 10%-90%, 20%-80%, 30%-70%…). Here we will show that when the probability was different from 50 percent, subjects overwhelmingly select the normatively superior option, i.e., Bag K if the probability of winning in this bag is 60 percent or higher, Bag U if the probability of winning in Bag K is 40 percent or lower. As a result, no clear differences between conditions can be detected in these scenarios.

Table 5.5: Probit Regression Results on Choosing Risk over Ambiguity

The table displays results from the Probit regression analyses of subjects' decisions to choose the risky Bag K (1) over the ambiguous Bag U (0). Definitions of the variables are as in previous tables. For each explanatory variable, the marginal effect is shown at the covariate means. Robust standard errors are used and *p*-values are shown in parentheses.

	Model 5.1		Model 5.2	
Condition dummies				
Eyes	-0.07	(0.46)	-0.11	(0.27)
Peers	-0.21	(0.02)	-0.19	(0.03)
Control variables				
Age			-0.03	(0.17)
Gender (female = 1)			0.11	(0.14)
Nationality (Dutch = 1)			0.06	(0.43)
Year of study				
Bachelor 2			0.06	(0.58)
Bachelor 3			0.08	(0.49)
Master			0.03	(0.84)
Calculator (yes = 1)			0.05	(0.49)
Log pseudo-likelihood		-85.79		-81.58
N		165		162

As discussed in the main chapter, the standard Ellsberg question we employed involved two bags containing black and red chips; in one bag (Bag K) the proportion of red and black chips was known, whereas in the second bag (Bag U) this proportion was not known. The subjects were asked to choose a color and a bag to draw a chip from. If the color of the drawn chip matched the one they had chosen, they received €50. When the proportion of red and black chips is 50-50, Bag K and Bag U are normatively equivalent, but many studies have shown that a disproportionate number of people choose Bag K (Camerer and Weber, 1992). In line with this common pattern, we observe that 85.5 percent of the subjects chose Bag K in our control condition. Using χ^2-tests, we find that behavior does not differ between the eyes and the control condition. In the peers condition, however, subjects are significantly less likely to show a bias in favor of Bag K.

Table 5.6: Descriptive Statistics of the Ambiguity Aversion Index

The table displays the descriptive statistics of the ambiguity aversion index over different conditions. We calculated the index by counting the number of times a subject prefers the risky prospect over all nine choice-tasks. The higher this index, the greater the degree to which the subject shows a preference for the risky prospect over the ambiguous one.

	N	Mean	Median	St.Dev.	min	max
Eyes	55	5,11	5,00	1,29	0	9
Peers	55	4,75	5,00	1,27	0	9
Control	55	5,02	5,00	0,91	2	7
Total	165	4,96	5,00	1,17	0	9

To investigate the robustness of this finding, we perform Probit analyses on the likelihood of choosing Bag K. The findings are reported in Table 5.5. As before, we report marginal effects around covariate means and apply robust standard errors in order to calculate statistical significance. These analyses yield results that are perfectly in line with the χ^2-tests reported in the paper. That is, subjects are significantly less likely to show a bias toward bag K in the peers condition as opposed to the control condition ($p < 0.03$). The difference between eyes and peers is marginally significant in Model 5.1 (p = 0.09, untabulated), but drops in significance when background characteristics are accounted for ($p = 0.32$, untabulated). The difference between the eyes and the control condition does not reach significance in any of the models ($p > 0.27$). None of the control variables influences the choice for Bag K. These analyses thus show the effect of the peers condition, as compared to the control condition, to be a rather robust phenomenon, whereas there is no evidence for an effect of the "eye" condition.

As mentioned above, we also asked subjects to choose between the ambiguous prospect and a range of risky prospects with a probability of winning of 10, 20, 30, 40, 60, 70, 80, and 90 percent. We start by using these questions to create an index of "ambiguity aversion", defined as the degree to which people tended to prefer the risky prospect to the ambiguous one. We generate this index by counting the number of times a subject prefers the risky prospect over all nine choice-tasks. The higher this index, the greater the degree to which the subject shows a preference for the risky prospect over the ambiguous one. This index indicates, as shown in Table 5.6, that subjects in the peers condition were less attracted by the risky urn. This difference is significant in Mann-Whitney tests (in comparison with the control: $p = 0.05$, comparison with the eyes condition: $p = 0.05$). The difference between the other two conditions is not significant ($p = 0.98$).

An interesting point, however, is that the above effect seems to be caused entirely by the choice when the probability of winning in the risky prospect is 50 percent. When we leave out this choice in our construction of the index, we find no significant differences between groups ($p > 0.21$). Figure 5.5 illustrates this point: when it comes to the index, the major differences arise around a score of four or five. Most subjects (87.3%) act consistently; they stick to the ambiguous prospect until the probability of winning in the risky prospects becomes sufficiently high, and after this point they consistently choose the risky prospects and do not switch back to the ambiguous one. Therefore, the switching points at which subjects decides to give up the ambiguous prospect for the risky ones drive the difference in the indexes that we observe between the conditions. Switching at the 50 percent risky prospect implies a score of five, switching prior to it at the 40 percent prospect implies a score of six, and switching only at the 60 percent prospect implies a score of four. Therefore, as can be clearly seen in Figure 5.5, the choice at the 50 percent prospect is the main driving force behind the differences in the ambiguity index.

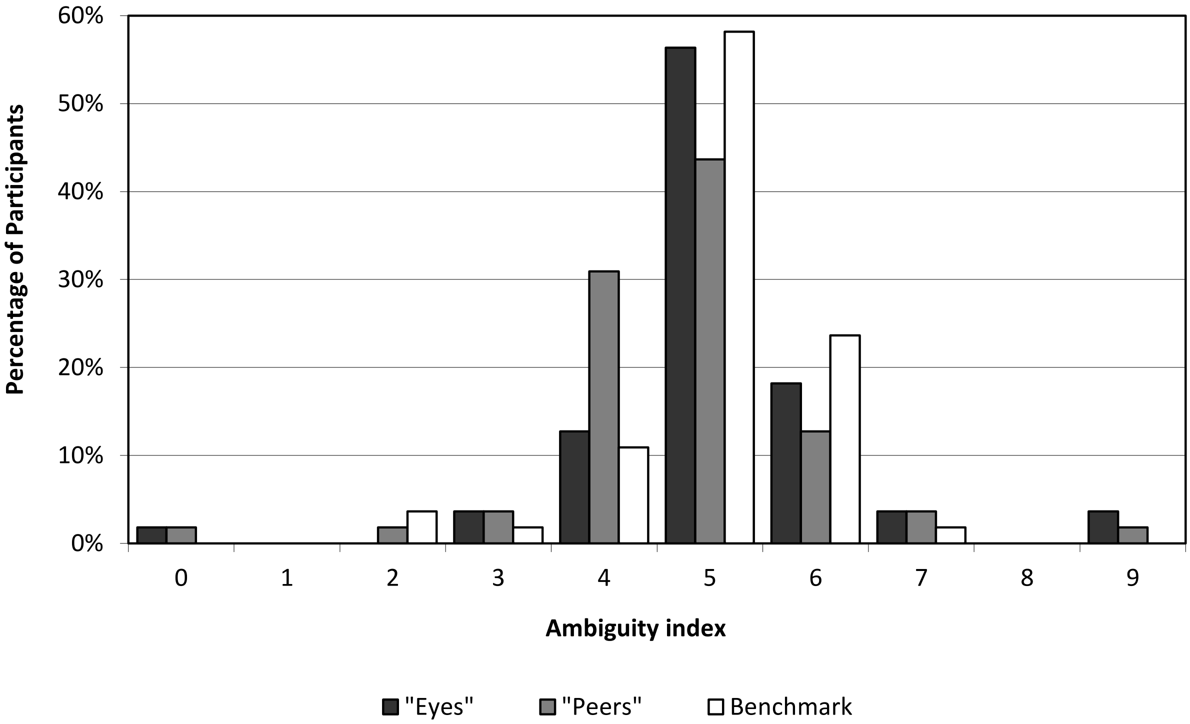


Figure 5.5: Ambiguity Index for the Three Conditions

The ambiguity aversion index is calculated by counting the number of times a subject prefers the risky prospect over all nine choice-tasks. The higher this index, the greater the degree to which the subject shows a preference for the risky prospect over the ambiguous one.

Finally, Figure S5 shows this finding by depicting the percentage of subjects who chose a risky prospect as a function of the probability of winning in that risky prospect. It is easy to see that when the probability of winning in the risky prospect is not 50 percent, most of the subjects show a strong preference for either of the two prospects: when the probability is lower than 50 percent, a strong majority of subjects choose the ambiguous prospect, and when it is higher than 50 percent, an overwhelming majority of subjects choose the risky one. Due to these strong majorities, we can no longer use *χ2*-tests to statistically test for differences between conditions in these tasks, as the *χ2*-test is not reliable when data is highly unbalanced. Therefore, we employ Fischer's exact test to test

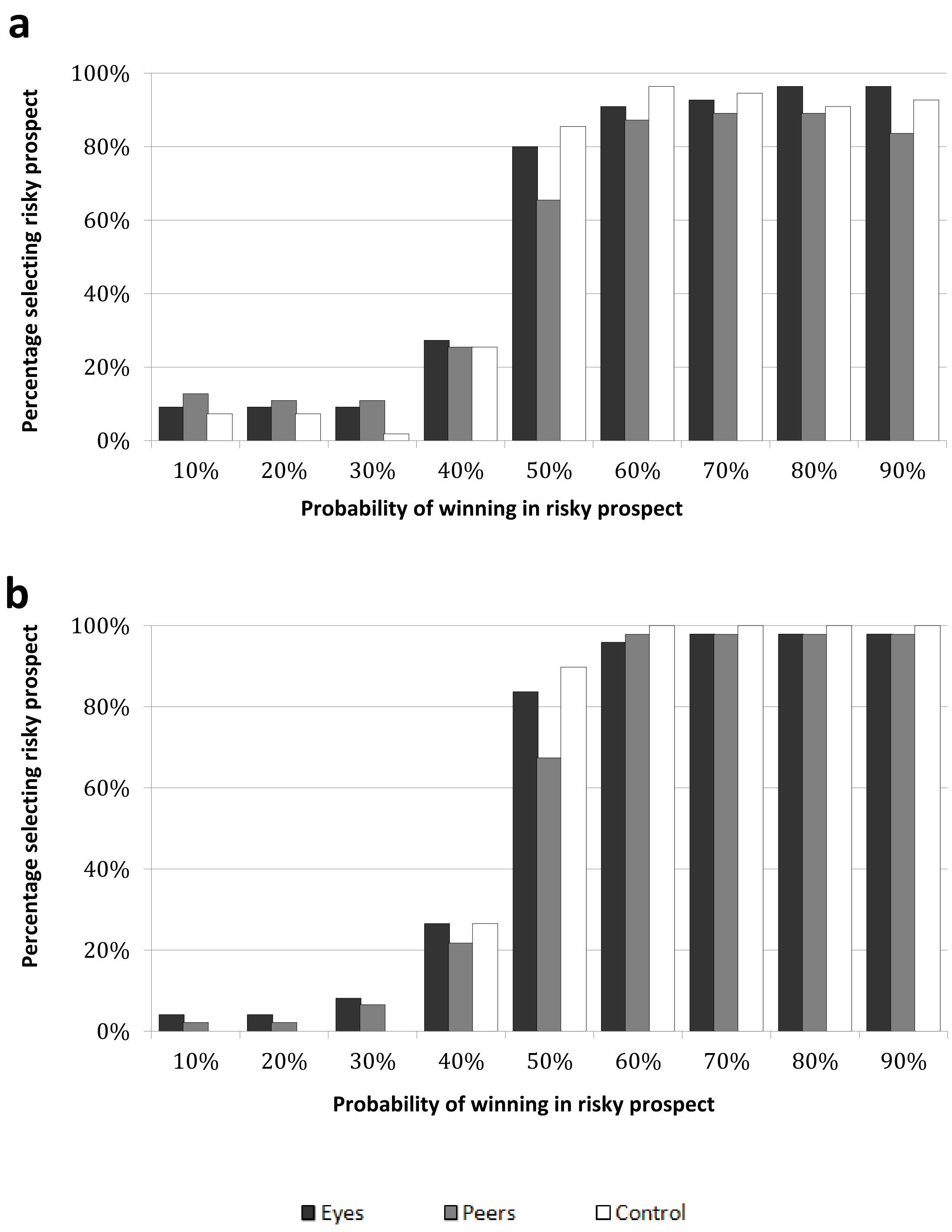


Figure 5.6: Choosing the Risky Prospect as Function of Winning Probability

The figures display the percentage of subject selecting the risky prospect for each of the nine choice questions that vary the probability of winning from 10% to 90% by condition. The Figure in (a) displays results for the complete set of data. The Figure in (b) displays results excluding a few subjects who showed inconsistent preferences in this task.

for differences between conditions in the choice task. It should be noted that applying Fischer's exact test does not alter our conclusions for choice regarding the 50 percent prospect, the difference between the peers and the control condition remains highly significant ($p = 0.03$), although the difference between the peers and the eyes condition is no longer significant ($p = 0.13$). For the other choice-tasks, neither the eyes nor the peers condition differs significantly from the control ($p > 0.11$). The only condition comparison that approaches significance is that between eyes and peers in the 90 percent choice task ($p = 0.05$).

It should, however, be noted that this difference completely disappears when we only focus on those 87.3 percent of the subjects who behave completely consistent within this task. If we leave the inconsistent subjects out of the analyses, the eyes and peers condition yield the exact same propensity to choose the risky prospect in the 90 percent choice task ($p = 1.00$). Focusing on these consistent individuals, again only the difference between the peers and the control condition at the 50 percent choice becomes significant ($p = 0.01$), while the difference between the eyes and peers condition becomes marginally significant ($p = 0.09$). No further condition differences emerge ($p > 0.11$). This implies that the only robust pattern in the ambiguity aversion task is the finding that when choosing between an ambiguous prospect and a risky prospect with a 50 percent winning probability, subjects in the peers condition are significantly less likely to show a bias in favor of the risky prospect as compared to the subjects in the control condition.

5.C.5 Simple vs. Compound Lotteries

As shown in the main chapter, we find that there are significant differences in the likelihood that subjects mistakenly choose the compound gamble over the superior simple gamble. While there is no difference between the eyes and the control condition, χ^2-tests indicates that the likelihood of making such a mistake is marginally significantly lower in the peers condition as compared to the other two conditions separately, and significantly lower if we combine the other two conditions. Furthermore, looking at the number of mistakes reveals a similar pattern. Mann-Whitney tests indicate that subjects in the peers condition make marginally significantly fewer errors in the peers as opposed to the control condition, where the eyes condition does not differ significantly from the other two conditions. In the present section we will show that these results are robust, or even strengthened, when performing more advanced analyses and controlling for the effect of other variables on the likelihood of making errors.

First, we perform a Probit analysis on the likelihood of making one or more mistakes. We consider four models: two in which we compare the eyes and peers condition to the control condition, and two in which we compare the peers condition to the other two. For both analyses we apply a simple model without control variables and a model that accounts for the effects of several control variables. Table 5.7 shows our results, again, using robust standard errors and reporting marginal effects evaluated at covariate means in order to give parameters a substantive meaning.

As Table 5.7 clearly shows the standard condition only models (Model 5.1 and Model 5.3), are perfectly in line with the χ^2-tests reported earlier; we observe no difference between the eyes and the control, a marginal significant difference between peers and the other two conditions separate (p = 0.08) and a significant difference between the peers condition and the two other conditions combined (p = 0.04). Adding the control variables (Model S7.2 and Model S7.4), we find that these results are robust and indeed increase in significance somewhat (respectively p = 0.06 and p = 0.01). Furthermore, females are more likely to make at least one error, Dutch student are less likely to do so, and the use of a calculator drastically decreases the likelihood of making an error.

Table 5.7: Probit Regression Results on the Likelihood of Making a Mistake

The table displays results from the Probit regression analyses on the likelihood that subjects choose a compound gamble over a strictly better simple gamble at least once. Model 5.1 and Model 5.2 compare both the eyes and the peers condition to the control, Model 5.3 and Model 5.4 compare the peers condition to the two other conditions combined. Definitions of the variables are as in previous tables. For each explanatory variable, we report marginal effects evaluated at covariate means. Robust standard errors are used and *p*-values are shown in parentheses.

	Control as reference				Control + eyes as reference			
	Model 5.1		Model 5.2		Model 5.3		Model 5.4	
Condition dummies								
Eyes	0.00	(1.00)	0.08	(0.42)				
Peers	-0.16	(0.08)	-0.23	(0.06)	-0.16	(0.04)	-0.27	(0.01)
Control variables								
Age			0.00	(0.87)			-0.01	(0.80)
Gender (female = 1)			0.19	(0.02)			0.19	(0.02)
Nationality (Dutch = 1)			-0.20	(0.02)			-0.19	(0.03)
Year of study								
Bachelor 2			0.05	(0.71)			0.05	(0.72)
Bachelor 3			-0.03	(0.87)			-0.02	(0.90)
Master			0.15	(0.41)			0.17	(0.33)
Calculator (yes = 1)			-0.58	(0.00)			-0.58	(0.00)
Log pseudo-likelihood		-107.66		-69.43		-107.66		-69.72
N		165		162		165		162

Secondly, we estimate an Ordinal Probit model where the dependent variable is the number of mistakes (0 through 6). Table S8 shows our results. We report coefficients with their significance levels. As in all previous analyses, significance levels are based on robust standard errors. Furthermore, we report marginal effects evaluated at covariate means for each possible outcome category (0 through 6 mistakes). We present two models: a basic condition model without control variables, and a full model, which includes control variables alongside the general condition effects.

Table 5.8: Ordinal Probit Regression Results on the Number of Mistakes

The table displays results from Ordinal Probit regression analyses on the number of mistakes that subjects make. The Condition model only includes condition dummies, whereas the Full model includes all our controls. We report both coefficients and marginal effects on the likelihood that a person makes a specific number of errors evaluated at covariate means. Robust standard errors are used and *p*-values are shown in parentheses.

	Coeff.		Marginal Effects						
			P(0)	P(1)	P(2)	P(3)	P(4)	P(5)	P(6)
Condition									
Condition dummies									
Eyes	-0.15	(0.48)	0.06	0.00	0.00	-0.01	-0.01	-0.02	-0.01
Peers	-0.43	(0.04)	0.16	0.00	-0.01	-0.03	-0.04	-0.05	-0.04
Log pseudo-likelihood	-285.94								
N	165								
Full									
Condition dummies									
Eyes	-0.10	(0.63)	0.04	0.00	0.00	-0.01	-0.01	-0.01	0.00
Peers	-0.52	(0.02)	0.20	0.01	-0.03	-0.06	-0.05	-0.04	-0.02
Control variables									
Age	-0.03	(0.66)	0.01	0.00	0.00	0.00	0.00	0.00	0.00
Gender (female = 1)	0.16	(0.40)	-0.06	-0.01	0.01	0.02	0.02	0.01	0.01
Nationality (Dutch = 1)	-0.56	(0.01)	0.20	0.02	-0.02	-0.06	-0.06	-0.05	-0.03
Year of study									
Bachelor 2	-0.01	(0.97)	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bachelor 3	0.16	(0.64)	-0.06	-0.01	0.01	0.02	0.02	0.01	0.01
Master	0.14	(0.72)	-0.05	0.00	0.00	0.02	0.01	0.01	0.01
Calculator (yes = 1)	-1.51	(0.00)	0.53	0.02	-0.06	-0.15	-0.14	-0.11	-0.08
Log pseudo-likelihood	-239.70								
N	162								

As can be seen in Table 5.8, both models show that the subjects in the peers condition are significantly less likely to make errors ($p < 0.04$). The difference between eyes and peers is insignificant in the first model ($p = 0.16$, untabulated) and marginally significant in the second model ($p = 0.06$, untabulated). With respect to the control variables, we find that Dutch students are significantly less likely to make mistakes, supporting the idea that this sub-group behaves more in line with rationality based arguments. Naturally, subjects who use calculators are also significantly less likely to make mistakes.

In conclusion, the non-parametric tests reported in the chapter are in line with the more advanced analyses including control variables reported here. In general, adding control variables seems to strengthen our results rather than weaken them.

Chapter 6 | Source-Dependence of Utility and Loss Aversion: A Critical Test of Ambiguity Models

This chapter tests whether utility is the same for risk and for uncertainty. This test is critical for models that capture ambiguity aversion through a difference in event weighting between risk and uncertainty, like the multiple priors models and prospect theory. We present a new method to measure utility and loss aversion under uncertainty without the need to introduce simplifying parametric assumptions. Our method extends Wakker and Deneffe's (1996) trade-off method by allowing for standard sequences that include gains, losses, and the reference point. It provides an efficient way to measure loss aversion and a useful tool for practical applications of ambiguity models. We cannot reject the hypothesis that utility and loss aversion are the same for risk and uncertainty, suggesting that utility primarily reflects attitudes towards outcomes. Utility is S-shaped, concave for gains and convex for losses and there is substantial loss aversion. Our findings support models that explain ambiguity aversion through a difference in event weighting and suggest that descriptive ambiguity models should allow for reference-dependence of utility.

This chapter is based on the paper "Source-Dependence of Utility and Loss Aversion: A Critical Test of Ambiguity Models", co-authored by Mohammed Abdellaoui, Han Bleichrodt, and Olivier L'Haridon (Abdellaoui *et al.*, 2013b). We gratefully acknowledge helpful comments from Aurélien Baillon, Ferdinand Vieider, Peter P. Wakker, and Horst Zank.

6.1 Introduction

An extensive amount of empirical work, originating from Ellsberg's (1961) famous thought experiment, shows that people are not neutral towards ambiguity, as assumed by subjective expected utility. New models have been proposed to explain these ambiguity attitudes. Broadly speaking, these ambiguity models can be subdivided into two classes. The first class models ambiguity aversion through a difference in utility between risk (known probabilities) and uncertainty (unknown probabilities). The best-known model of this class is the smooth ambiguity model of Klibanoff, Marinacci, and Mukerji (2005). Other models that belong to this class were proposed by Nau (2006), Chew *et al.* (2008), Seo (2009), and Neilson (2010). The second class of models assumes that utility does not depend on the source of uncertainty and is the same for risk and uncertainty. Instead, ambiguity aversion is modeled through a difference in event weighting. This class includes the multiple priors models (Gilboa and Schmeidler, 1989; Jaffray, 1989; Ghirardato, Maccheroni, and Marinacci, 2004) and modifications thereof (Gajdos *et al.,* 2008; Maccheroni, Marinacci, and Rustichini, 2006), vector expected utility (Siniscalchi, 2009), Choquet expected utility (Gilboa, 1987; Schmeidler, 1989), and prospect theory (Kahneman and Tversky, 1979; Tversky and Kahneman, 1992).

This chapter investigates whether utility is source-independent and the same for risk and uncertainty. We assume a general utility model, previously suggested by Miyamoto (1988), Luce (1991), and Ghirardato and Marinacci (2001), that includes most of the ambiguity models of the second class as special cases, and generalize it to include sign-dependence to also cover prospect theory. We test the central condition underlying this model and obtain support for it. We measure utility for gains and for losses and also measure loss aversion. Previous evidence suggests that the distinction between gains and losses is relevant because ambiguity attitudes differ between gains and losses (*e.g.,* Cohen, Jaffray, and Said, 1987; Hogarth and Kunreuther, 1989; Abdellaoui, Vossmann, and Weber, 2005; Du and Budescu, 2005) and loss aversion is crucial in explaining attitudes towards both risk (Rabin 2000) and ambiguity (Roca, Hogarth, and Maule, 2006).

Measuring loss aversion is complex, in particular if event weighting may be different for gains and losses. Previous measurements of loss aversion sidestepped this problem by introducing simplifying assumptions. We introduce a new method to measure loss aversion that imposes no simplifying assumptions and requires no complete measurement of utility. It can easily be applied, which may encourage the use of ambiguity models in decision analysis. Our method extends the trade-off method of Wakker and Deneffe (1996) by allowing standard sequences (sequences of outcomes for which the utility difference between successive elements is constant) to pass through the reference point. Our method also simplifies the axiomatization of ambiguity models as there is a close connection between measurements of utility using the trade-off method and preference conditions (Köbberling and Wakker, 2003).

Our experimental data contain two messages. First, they provide support for models that explain ambiguity aversion through a difference in event weighting. We cannot reject the hypothesis that utility and loss aversion are the same for risk and uncertainty. This suggests that utility is source-independent and primarily reflects attitudes towards outcomes.

The second message is that descriptive ambiguity models should allow for reference-dependence of utility. We obtain clear evidence that utility differs for gains and losses and there was sizeable loss aversion. Most ambiguity models do not allow for reference-dependence and assume that ambiguity attitudes are the same for gains and losses. This assumption may be adequate for normative purposes, but, as our data clearly show, does not match behavior.

6.2 Background

Binary Prospect Theory

Consider a decision maker who has to make a choice in the face of uncertainty. Uncertainty is modeled through a *state space* S. Exactly one of the states will obtain, but the decision maker does not know which one. Subsets E of S are called *events* and E^c denotes the complement of E.

Acts map states to outcomes. Outcomes are money amounts and more money is preferred to less. In our measurements, we will only use two-outcome acts $x_E y$, signifying that the decision maker obtains € x if event E occurs and € y otherwise. If probabilities are known, we will write $x_p y$ for the act that pays € x with probability p and € y with probability $1-p$. We will refer to $x_E y$ as an *uncertain act* (meaning that probabilities are unknown) and to $x_p y$ as a *risky act* (meaning that probabilities are known).

We use conventional notation to express the preference of the decision maker, letting $\succ$, $\succcurlyeq$, and $\backsim$ represent strict preference, weak preference, and indifference. Preferences are defined relative to a reference point x_0. *Gains* are outcomes strictly preferred to x_0 and *losses* are outcomes strictly less preferred than x_0. An act is *mixed* if it involves both a gain and a loss. For mixed acts the notation $x_E y$ signifies that x is a gain and y is a loss. A *gain act involves no losses (i.e. both x and y are nonnegative) and a loss act involves* no gains. For gain and loss acts the notation $x_E y$ signifies that the absolute value of x exceeds the absolute value of y, *i.e.* if x and y are gains then $x \geq y$ and if x and y are losses then $x \leq y$.

Under *binary prospect theory (PT)* the decision maker's preferences over mixed acts $x_E y$ are evaluated by:

(6.1a) $$W^+(E)U(x) + W^-(E^c)U(y)$$

and preferences over gain or loss acts by:

(6.1b) $$W^i(E)U(x) + \left(1 - W^i(E)\right)U(y)$$

where $i = +$ for gains and $i = -$ for losses. U is a strictly increasing, real-valued *utility function* that satisfies $U(x_0) = 0$. The utility function is a ratio scale and we can choose the utility of one outcome other than the reference point. U is an overall utility function that includes loss aversion. In empirical applications U is often decomposed in a basic utility function, capturing the decision maker's attitudes towards final outcomes, and a loss aversion coefficient λ capturing attitudes towards gains and losses (Sugden, 2003; Köbberling and Wakker, 2005; Köszegi and Rabin, 2006). Our method does not require this decomposition.

The *event weighting functions* $W^i, i = +, -$, assign a number $W^i(E)$ to each event E such that

$$W^i(\emptyset) = 0$$

$$W^i(S) = 1$$

W^i is *monotonic*: $E \supset F$ implies $W^i(E) \geq W^i(F)$.

The event weighting functions W^i depend on the sign of the outcomes and may be different for gains and losses. They need not be additive. For gains, binary PT contains most transitive ambiguity models as special cases, as was pointed out by Miyamoto (1988), Luce (1991), and Ghirardato and Marinacci (2001). The ambiguity models only differ when the number of outcomes is at least three. Equations (6.1a) and (6.1b) represent the extension of these models to include sign-dependence.

Binary PT evaluates mixed risky acts $x_p y$ as

$$\text{(6.2a)} \qquad w^+(p)U(x) + w^-(1-p)U(y)$$

and gain and loss risky acts $x_p y$ as

$$\text{(6.2b)} \qquad w^i(p)U(x) + \left(1 - w^i(p)\right)U(y), i = +, -.$$

w^i is a strictly increasing *probability weighting function* that satisfies $w^i(0) = 0$ and $w^i(1) = 1$ and again may differ between gains and losses. Hence, in the evaluation of risky acts the event weighting functions W^i are replaced by probability weighting functions w^i. Binary PT assumes that utility is the same for risk and uncertainty. Ambiguity aversion is modelled through a difference between W^i and w^i.

Previous Evidence

Tversky and Kahneman (1992) assumed that utility differs between gains and losses and is S-shaped, concave for gains and convex for losses. In addition, they assumed that utility is steeper for losses than for gains, reflecting loss aversion. Nearly all the empirical evidence on utility comes from decision under risk. There is much evidence that utility for gains is indeed concave (Wakker, 2010). For losses the evidence is more equivocal. While most studies found convex utility, some have also found linear or concave utility (*e.g.,* Bruhin, Fehr-Duda, and Epper, 2010). The utility for losses is usually closer to linear than the utility for gains.

Empirical evidence on utility under uncertainty is scarce. Abdellaoui, Vossmann, and Weber (2005) measured utility under uncertainty and confirmed that it was concave for gains and slightly convex for losses. Their parametric estimates were close to those previously obtained under risk, but they did not directly measure utility under risk. Abdellaoui *et al.* (2011) and Vieider *et al.* (2013) measured utility under risk and under uncertainty for small stakes and under parametric assumptions about utility. They found that utility was linear both for risk and for uncertainty. This finding might be due to the small stakes used in these studies: for small stakes utility is usually close to linear (Wakker, 2010).

Nearly all empirical measurements of loss aversion made simplifying assumptions about utility and probability weighting, typically assuming linear utility and either ignoring probability weighting (Pennings and Smidts, 2003; Booij and van de Kuilen, 2009; Baltussen, van den Assem, and van Dolder*,* 2013) or assuming equal weighting for gains and losses (Gächter, Johnson, and Herrmann, 2010). The exception is Abdellaoui, Bleichrodt, and Paraschiv (2007) who imposed no simplifying assumptions on either probability weighting or utility. However, they measured loss aversion in decision under risk only and their method is not applicable in decision under uncertainty.

Most studies found loss aversion coefficients around 2, meaning that losses weight approximately twice as much as absolutely commensurate gains (Booij, van Praag, and van de Kuilen, 2010). A difficulty in comparing the results of these studies is that they not only made different parametric assumptions, but also adopted different definitions of loss aversion.

Finally, even though binary PT is consistent with much of the empirical data that has been collected on decision under risk and uncertainty and includes many ambiguity models as special cases, there is some evidence challenging it. For example, Starmer and Sugden (1993) and Birnbaum (2008) reported event-splitting effects that violate binary PT and Birnbaum and Bahra (2007) and Wu and Markle (2008) obtained violations of binary PT for mixed acts. We, therefore, included a test of the main condition underlying binary PT in our experiment. This test is explained below.

6.3 Measurement Method

Our method for measuring utility and loss aversion consists of three stages and is summarized in Table 6.1. In the first stage, a gain and a loss are elicited that connect utility for gains (measured in the second stage) with utility for losses (measured in the third stage). The measurements in the second and in the third stage employ the trade-off method of Wakker and Deneffe (1996). Within each domain, we determine a standard sequence of outcomes such that the utility difference between successive elements of the sequence is constant. The trade-off method is commonly used in decision theory (Wakker, 2010), but thus far it could only be used to measure utility for gains and utility for losses separately. It could not be used to measure loss aversion, which requires that the utility for gains and the utility for losses can be compared. Our method allows measuring utility for gains and utility for losses jointly and, consequently, it permits the measurement of loss aversion. In all the derivations presented below we impose no parametric assumptions on utility and the weighting functions W^i and $w^i, i = +, -,$. Hence, our method is parameter-free.

Table 6.1: Three-Stage Procedure to Measure Utility

The third column shows the quantity that is assessed in each of the three stages of the procedure. The fourth column shows the indifference that is elicited. The fifth column shows the stimuli used in the experiment. ℓ_{alt} and $k_{L_{alt}}$ were used to test for consistency (see Section 6.4 for explanation).

		Assessed quantity	Indifference	Choice variables
		L	$G_E L \sim x_0$	$G = €2000$
Stage 1		x_1^+	$x_1^+ \sim G_E x_0$	E = color of a ball drawn from an
		x_1^-	$x_1^- \sim L_{E^c} x_0$	unknown Ellsberg urn, $p = ½$
Stage 2	Step 1	$\mathcal{L}$	$x_1^+ {}_E \mathcal{L} \sim \ell_{E^c} x_0$	$\ell = -€300$; $k_G = 6$
	Step 2 to k_G	x_j^+	$x_j^+ {}_E \mathcal{L} \sim x_{j-1}^+ {}_E \ell$	$\ell_{alt} = €0$; $k_{G_{alt}} = 3$
Stage 3	Step 1	$\mathcal{G}$	$\mathcal{G}_E x_1^- \sim g_E x_0$	$g = €300$; $k_L = 6$
	Step 2 to k_L	x_j^-	$\mathcal{G}_E x_j^- \sim g_E x_{j-1}^-$	

First Stage: Elicitation of the Gauge Outcomes

We start by selecting an event E that will be kept constant throughout the first stage and a gain G. Then we elicit the loss L for which $G_E L \sim x_0$. It follows from equation (6.1a) that:

(6.3) $$W^+(E)U(G) + W^-(E^c)U(L) = U(x_0) = 0$$

We next elicit certainty equivalents x_1^+ and x_1^- such that $x_1^+ \sim G_E x_0$ and $x_1^- \sim L_{E^c} x_0$. The indifference $x_1^+ \sim G_E x_0$ implies that

(6.4) $$U(x_1^+) = W^+(E)U(G).$$

The indifference $x_1^- \sim L_{E^c} x_0$ implies that

(6.5) $$U(x_1^-) = W^-(E^c)U(L).$$

Combining Eqs. (6.3)– (6.5) gives

(6.6) $$U(x_1^+) = -U(x_1^-).$$

Equation (6.6) defines the first elements x_1^+ and x_1^- of the standard sequences for gains and losses that we will construct in the second and third stages.

For choice under risk, the elicitation of x_1^+ and x_1^- is similar except that the event E is replaced by a known probability p, and that the weights $W^+(E)$ and $W^-(E^c)$ are replaced by $w^+(p)$ and $w^-(1-p)$, respectively.

Second and Third Stage: Elicitation of Utility for Gains and Losses

In the second stage, we elicit a standard sequence of gains. Let ℓ be a prespecified loss. We first elicit the loss $\mathcal{L}$ such that the decision maker is indifferent between the acts $x_1^+{}_E\mathcal{L}$ and $\ell_{E^c}x_0$, where x_1^+ is the gain that was elicited in the first stage. We can take an event E' different from the event E used in the first stage, but, for simplicity, we applied the same event in all three stages in our experiment. The indifference $x_1^+{}_E\mathcal{L}\sim\ell_{E^c}x_0$ implies that

$$(6.7) \qquad W^+(E)U(x_1^+) + W^-(E^c)U(\mathcal{L}) = W^-(E^c)U(\ell).$$

Rearranging Eq. (6.7) and using $U(x_0) = 0$ gives,

$$(6.8) \qquad U(x_1^+) - U(x_0) = \frac{W^-(E^c)}{W^+(E)}\big(U(\ell) - U(\mathcal{L})\big).$$

Next, we elicit the gain x_2^+ such that $x_2^+{}_E\mathcal{L}\sim x_1^+{}_E\ell$. From this indifference we obtain after rearranging

$$(6.9) \qquad U(x_2^+) - U(x_1^+) = \frac{W^-(E^c)}{W^+(E)}\big(U(\ell) - U(\mathcal{L})\big).$$

Combining Eqs. (6.8) and (6.9) gives:

$$(6.10) \qquad U(x_2^+) - U(x_1^+) = U(x_1^+) - U(x_0).$$

We proceed by eliciting a series of indifferences $x_j^+{}_E\mathcal{L}\sim x_{j-1}^+{}_E\ell, j = 2, \dots, k_G$, to obtain the sequence $\{x_0, x_1^+, x_2^+, \dots, x_{k_G}^+\}$. It is easy to see that for all j, $U(x_j^+) - U(x_{j-1}^+) = U(x_1^+) - U(x_0)$. For decision under risk, we apply the above procedure with the event E replaced by a probability p.

The standard sequence of losses is constructed similarly. We select a gain g and an event E and elicit the gain $\mathcal{G}$ such that $\mathcal{G}_E x_1^- \sim g_E x_0$.[45] We then proceed to elicit a standard

[45] Again, we could have selected an event E" different from the events used in the first two stages, but we used the same event in our experiment.

sequence $\{x_0, x_1^-, x_2^-, \dots, x_{k_L}^-\}$ by eliciting a series of indifferences $\mathcal{G}_E x_j^- \sim \mathcal{g}_E x_{j-1}^-$, $j = 2, \dots, k_L$. For risk, we replace the event E by a probability p.

By combining the second and the third stages we elicit a sequence $\{x_{k_L}^-, \dots, x_1^-, x_0, x_1^+, \dots, x_{k_G}^+\}$ that runs from the domain of losses through the reference point to the domain of gains and for which the utility difference between successive elements is constant. We can scale utility by selecting the utility of an arbitrary element. In the analyses reported below, we set $U(x_{k_G}^+) = 1$ from which it follows that $U(x_j^+) = j/k_G$ for $j = 1, \dots, k_G$, and $U(x_j^-) = -j/k_G$, for $j = 1, \dots, k_L$.

6.4 Experiment

Experimental Set-Up

Subjects were 75 economics students of the Erasmus School of Economics, Rotterdam (29 female, mean age of 20.7 years). Each subject was paid a flat fee of €10 for participation in the experiment. Before conducting the actual experiment, the experimental protocol was tested in several pilot sessions.

The experiment was run on computers. Subjects answered the questions individually in sessions of at most two subjects. They first received instructions about the tasks and then completed five training questions. Subjects were told that there were no right or wrong answers and that they should go through the experiment at their own pace. They were instructed to approach the experimenter if they needed any advice concerning the experiment. A session lasted 40 minutes on average.

The order in which utility under risk and uncertainty were measured was randomized between sessions. When a subject had completed the first part of the experiment, the experimenter would approach her to explain the next part. Within the risk and uncertainty elicitations, the second and third stage were also randomized; some subjects started with the elicitation of the gain sequence, others with the elicitation of the loss sequence. The first stage always had to come first because it served as an input for the other two stages.

We used sizeable monetary amounts because we were interested in studying both utility curvature and loss aversion. Utility is approximately linear over small intervals (Wakker and Deneffe, 1996) and we feared that it would be hard to detect differences between utility under risk and uncertainty for small stakes. Given that substantial losses were involved, all choices were hypothetical. It is impossible to find subjects willing to participate in an experiment where they can lose substantial amounts of money. We will provide a more detailed discussion of the use of incentives in the concluding section.

We did not directly ask subjects for their indifference values, but, instead, used a series of binary choice questions to zoom in at them. Examples of such a zooming-in process can be found in

Table 6.5 in Appendix 6.B. We applied a choice-based elicitation procedure as previous research suggests that it leads to more reliable results than directly asking for indifference values (Bostic, Herrnstein, and Luce, 1990; Noussair, Robin, and Ruffieux, 2004).

Details

To perform the elicitation described in Section 6.3, we had to specify a number of parameters, which are depicted in the final column of Table 6.1. We made the common assumption that the reference point x_0 was equal to 0. In the risk condition, the outcome of an act was determined by drawing a ball from an urn containing five red balls and five black balls. Subjects could state which color they preferred to bet on with the chance of winning always equal to 50 percent. In the uncertainty condition, the outcome of an act was determined by drawing a ball from an urn containing ten balls, which were either red or black in unknown proportions. Again, subjects could select the color they preferred to bet on.

Both for gains and for losses, we elicited six points of the utility function under both risk and uncertainty. Next to these elicitations, we performed a second smaller sequence in the domain of gains, varying the gauge amount ℓ. By definition ℓ needs to be smaller or equal to x_0. In the main elicitation we set $\ell = -€300$. Asking the question whether the elicited amounts would depend on the value of ℓ, we also elicited x_2^+ and x_3^+ using an alternative gauge amount $\ell_{alt} = €\, 0$. Under binary PT the elicitations of x_2^+ and x_3^+ should not depend on the selected value of ℓ. This second elicitation was meant to test sign-comonotonic trade-off consistency (Köbberling and Wakker, 2003), the central condition underlying binary PT.

Figures 6.4-6.6 in the appendix 6.A show the displays used under uncertainty. The screens under risk were similar, except that the two branches would simply say 50% rather than "Red" or "Black". Figure 6.6 displays the typical decision that subject had to make. Subjects were faced with a choice between two acts denoted as options A and B. They could not state indifference. By choosing between the two acts, the subject narrowed down the interval in which her indifference value should fall.

After narrowing down the interval thrice, we presented subjects with a scrollbar (Figure 6.6). The scrollbar allowed subjects to specify their indifference value up to €1 precision. The starting point of the scrollbar was in the middle of the interval determined by their previous choices. The range of the scrollbar was wider than this interval, so that subjects could correct any mistakes they might have made. The data on the use of the scrollbar also give an indication of the quality of the data. If many subjects would provide answers that did not align with their previous choices, possibly even violating stochastic dominance, this might signal poor understanding of the task. After specifying a value with the scrollbar, subjects were asked to confirm their choice (Figure 6.7). If they cancelled their choice, the process started over. If subjects confirmed their choice, they moved on to the next elicitation.

We included a number of repetitions to test for consistency. First, in each of the six standard sequences (the short and the long gain sequences and the loss sequence for both risk and uncertainty), we repeated the second-to-last iteration in the elicitation of $x_2^i, i = +, -$. Repeating the second-to-last iteration is a strong test of consistency, as subjects were probably close to indifference at the end of the iteration process. Furthermore, at the end of eliciting the long gain sequence, we elicited x_4^+ again, for both risk and uncertainty. Together, these repetitions and the way in which subjects used the scrollbar allowed us to gain insight into the quality of the data.

6.5 Analyses

Analyses of Utility Curvature

We employ two different methods to investigate utility curvature.[46] For the first, nonparametric, method, we calculate the area under the utility function. The domain of U is normalized to $[0,1]$, by transforming every gain x_j^+ to the value x_j^+/x_6^+ and every loss x_j^- to x_j^-/x_6^-.[47] If utility is linear, the area under this normalized curve equals ½. For gains, we consider utility to be convex [concave] if the area under the curve is smaller [larger] than ½. For losses, utility is considered to be convex [concave] if the area under the curve is larger [smaller] than ½.

We also analyze the utility function by parametric estimation. We employ the power family, x^α, as it is the most commonly employed parametric family. For gains [losses] $\alpha > 1$ corresponds to convex [concave] utility, $\alpha = 1$ corresponds to linear utility, and $\alpha < 1$ corresponds to concave [convex] utility. Estimation is done using nonlinear least squares. To test for robustness, we have also performed a mixed-effects estimation in which each individual parameter was estimated as the sum of a fixed effect, common to all subjects, and an individual-specific random effect. The results were similar. A potential problem in estimating a model like binary PT using nonlinear least squares is collinearity between utility and the event weights, which implies that the obtained estimates may not be uniquely identified. The trade-off method avoids this problem by keeping event weighting fixed, while eliciting utility and, hence, the obtained estimates are uniquely identified.

Loss Aversion

In the literature, loss aversion has been defined in a multitude of ways. Abdellaoui, Bleichrodt, and Paraschiv (2007) concluded that the definitions proposed by Kahneman and

[46] We have also used a third, nonparametric, method based on changes in the slope of utility. This method leads to similar conclusions.

[47] One subject violated monotonicity so that x_6^- was not the largest loss. For this subject we transform losses x_j^- to $x_j^-/\{\min_{i=1,\dots,6} x_i^-\}$.

Tversky (1979) and Köbberling and Wakker (2005) were empirically most useful, and we will use these. Other definitions (Wakker and Tversky, 1993; Bowman, Minehart, and Rabin, 1999; Neilson, 2002) turned out to be too strict for empirical purposes, leaving many subjects unclassified.

Kahneman and Tversky (1979) defined loss aversion as $-U(-x) > U(x)$ for all $x > 0$. To measure loss aversion, we compute $-U\left(-x_j^+\right)/U\left(x_j^+\right)$ and $-U\left(-x_j^-\right)/U\left(x_j^-\right)$ for $j = 1, \dots, 6$, whenever possible.[48] Usually $U(-x_j^+)$ and $U(-x_j^-)$ cannot be observed directly and has to be determined through linear interpolation. Some subjects occasionally violate stochastic dominance. As a result, their utility is not unique and one amount can have multiple utilities. For these amounts, we consider utility to be undefined. A subject is classified as loss averse if $-U(-x)/U(x) > 1$ for all observations, as loss neutral if $-U(-x)/U(x) = 1$ for all observations, and as gain seeking if $-U(-x)/U(x) < 1$ for all observations. To account for response error, we also use a more lenient approach, classifying subjects as loss averse, loss neutral, or gain seeking if the above inequalities hold for more than half of the observations.

Köbberling and Wakker (2005) defined loss aversion as the kink of utility at the reference point (Bernatzi and Thaler (1995) suggested a similar definition). Formally, they defined loss aversion as $U'_{\uparrow}(0)/U'_{\downarrow}(0)$, where $U'_{\uparrow}(0)$ represents the left derivative and $U'_{\downarrow}(0)$ the right derivative of U at the reference point. To operationalize this empirically, we compute each subject's coefficient of loss aversion as the ratio of $U(x_1^-)/x_1^-$ over $U(x_1^+)/x_1^+$, because x_1^- and x_1^+ are the loss and gain closest to the reference point. Given that $U(x_1^-) = -U(x_1^+)$, this ratio is equal to $x_1^+/-x_1^-$. Hence, our method immediately gives an approximation of Köbberling and Wakker's (2005) loss aversion coefficient without the need to further measure utility. A subject is classified as loss averse if this ratio exceeds 1, as loss neutral of it is equal to 1, and as gain seeking if it is smaller than 1.

[48] These computations require that $-x_j^+$ is contained in $[x_6^-, 0)$ and $-x_j^-$ in $(0, x_6^+]$.

6.6 Results

Three subjects violated stochastic dominance in the first stage of the measurement procedure. This undermines their subsequent answers and they are removed from the analyses. For the remaining 72 subjects, we can determine the entire utility function, for both gains and losses and under both risk and uncertainty. Of these 72 subjects, 14 violated stochastic dominance at least once. Violations of stochastic dominance potentially signal a lower degree of understanding or a lower degree of effort put in the task. We have, therefore, also analyzed the data including only the 58 subjects who never violated stochastic dominance, leading to similar conclusions.

Consistency Checks

Overall, consistency is satisfactory. Subjects make the same choice in 63.7 percent of the repeated choices. Reversal rates round ⅓ are common in the literature (Stott, 2006). Moreover, our consistency test is strict, as subjects are close to indifference in the repeated choice and, hence, reversals are more likely. There is no difference in consistency between risk and uncertainty.

The correlation between the original measurement and the repeated measurement of x_4^+ is almost perfect. For risk, Kendall's τ is 0.924, for uncertainty it is 0.938.

As a final indication of consistency, we compare whether the final answer provided by using the slider fell within the interval as set up by the bisection procedure. Subjects provided answers that align with their original choices. Furthermore, when a subject's final answer is outside the bisection interval, it typically only violated the final choice, probably indicating that they were close to indifference at this point.

A Test of Binary PT

As explained in Section 6.4, we elicited two sequences of gains, a longer one based on $\ell = -€300$, which we use in the main analysis, and a shorter one based on $\ell_{alt} = €0$. If our subjects behave according to binary PT, then the values of x_2^+ and x_3^+ in the short sequence should be equal to those obtained in the long sequence.

We find support for binary PT, both for risk and for uncertainty. The correlation between the obtained values is substantial. Under risk, Kendall's τ is 0.564 for x_2^+ and 0.518 for x_3^+. Under uncertainty, these values are 0.694 for x_2^+ and 0.625 for x_3^+. All correlation coefficients are different from 0 ($p < 0.001$). Moreover, for uncertainty, we cannot reject the hypotheses that the values of x_2^+ and x_3^+ obtained in the short sequence are equal to those obtained in the long sequence (Mann-Whitney *U* test: both $p > 0.684$). For risk, the values of x_2^+ differ marginally ($p = 0.055$), but the values of x_3^+ do not differ ($p = 0.138$). Hence, even though x_3^+ is chained to x_2^+, the marginal difference for x_2^+ does not carry over to x_3^+.

Ambiguity Aversion

The measurement of L and x_1^+ in stage 1 of our method provide insight into subjects' ambiguity attitudes. Let L_r and L_u denote the elicited values of L for risk and uncertainty, respectively. Then, $2000_{.5}L_r \sim 0$ and $2000_E L_u \sim 0$. A subject is ambiguity averse if $2000_{.5}L_r \succ 2000_E L_r$. By transitivity, $2000_E L_u \succ 2000_E L_r$ and, thus, $L_u > L_r$. This is true for 63.9 percent of our subjects (Binomial test: $p = 0.024$) and the median elicited value of L_u ($-€612.50$) indeed exceeds the median value of L_r ($-€750$) (Mann-Whitney *U* test: $p = 0.012$). Hence, we find evidence of ambiguity aversion in the measurement of L.

Ambiguity aversion also predicts that $x_{1,r}^+$, the value of x_1^+ measured under risk will exceed $x_{1,u}^+$, the value of x_1^+ measured under uncertainty. This follows by transitivity from $x_{1,r}^+ \sim 2000_{.5}0 \succ 2000_E 0 \sim x_{1,u}^+$. However, this is only true for 44.4 percent of our subjects and we cannot reject the hypothesis that x_1^+ was the same for risk and for uncertainty (Mann-Whitney *U* test: $p = 0.807$).

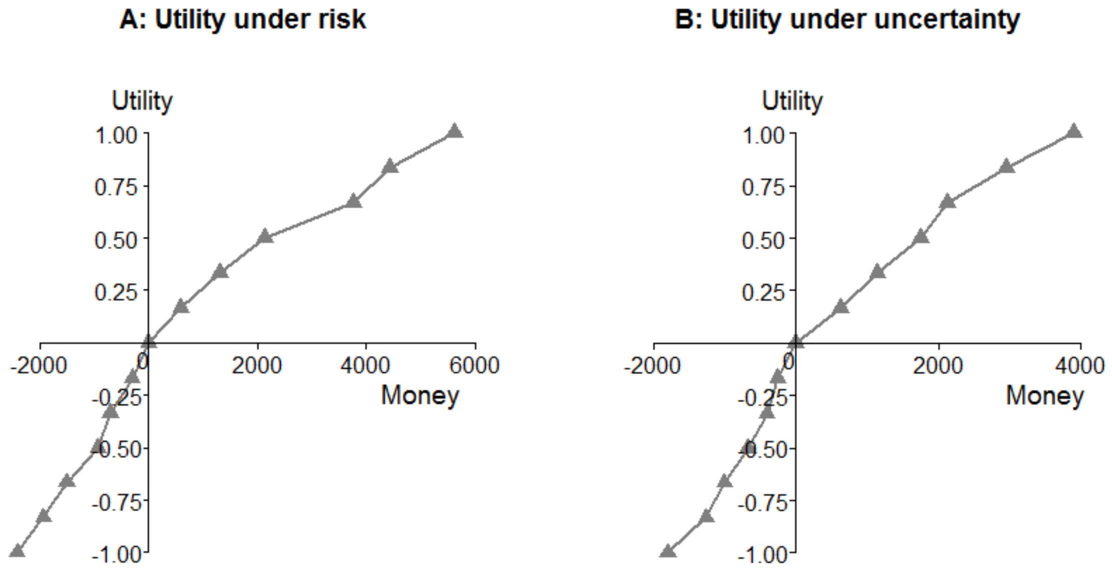


Figure 6.1: Utility for Gains and Losses Under Prospect Theory Based on Median Data

The figure displays the utility for gains and losses under prospect theory based on the median responses of our subjects. Panel A displays utility under risk. Panel B displays utility under uncertainty.

The Utility for Gains and Losses

Figure 6.1, Panel A displays the utility for gains and losses under risk, based on the median data. Figure 6.1, Panel B shows the same graph for uncertainty. Taken at face value, the utility functions seem similar. They are consistent with the typical finding of convex utility for losses and concave utility for gains. Furthermore, both utility functions appear considerably steeper for losses than for gains, indicating loss aversion.

Table 6.2: Classification of Subjects According to the Shape of their Utility Function

The table classifies the subjects according to the shape of their utility function based on the area under the normalized utility function. Panel A displays the results under risk. Panel B displays the results under uncertainty.

Panel A: Risk				
	Losses			
Gains	Concave	Convex	Linear	Total
Concave	13	31	1	45
Convex	15	8	1	24
Linear	2	0	1	3
Total	30	39	3	72
Panel B: Uncertainty				
	Losses			
Gains	Concave	Convex	Linear	Total
Concave	13	30	0	43
Convex	18	10	2	30
Linear	1	0	0	1
Total	32	40	2	72

To investigate these patterns more thoroughly, we move to the individual level analysis. Table 6.2 shows that the classification of subjects according to the shape of their utility function is very similar for risk and uncertainty and there are no differences in the overall distribution of classifications between conditions (Fisher's exact test: p = 0.943). Utility under risk and uncertainty are related (Kendall's τ = 0.389 for gains and 0.455 for losses, p < 0.001 in both cases) and the common pattern is that of an S-shaped utility function: concave for gains and convex for losses. Less than 20% of the subjects behave according to the traditional assumption in decision theory that utility is concave throughout.

The parametric results confirm the above conclusions. Table 6.3 shows the estimated power functions at the individual level. Utility is mostly concave for gains and convex for losses. Under risk, 32 subjects have S-shaped utility. Under uncertainty, this is the case for 30 subjects.

Table 6.3: Summary of Individual Power Coefficients for Gains and Losses

The table depicts the results of fitting power functions on each subject's choices individually. Shown are the m edian and interquartile range (IQR) for the resulting estimates.

	Risk		Uncertainty	
	Gains	Losses	Gains	Losses
Median	0.857	0.924	0.937	0.898
IQR	[0.616-1.062]	[0.649-1.154]	[0.716-1.188]	[0.675-1.356]

Figure 6.2 shows the relationship between individual estimates for the power coefficients under risk and uncertainty. The dashed lines correspond to the case where subjects have exactly the same coefficients in the two domains. Most estimates are relatively close to the dashed lines and there is no strong indication that subjects had different curvature under risk than under uncertainty.

Mann-Whitney *U* tests on these power function estimates indicate that there is no difference in curvature for losses between risk and uncertainty (p = 0.866). There is some indication that utility for gains is more concave under risk (p = 0.027).[49] The power coefficients of utility under risk and under uncertainty are moderately correlated: Kendall's τ being 0.373 for gains and 0.423 for losses.

[49] The difference is not significant if we restrict our attention to the 58 subjects who never violate stochastic dominance.

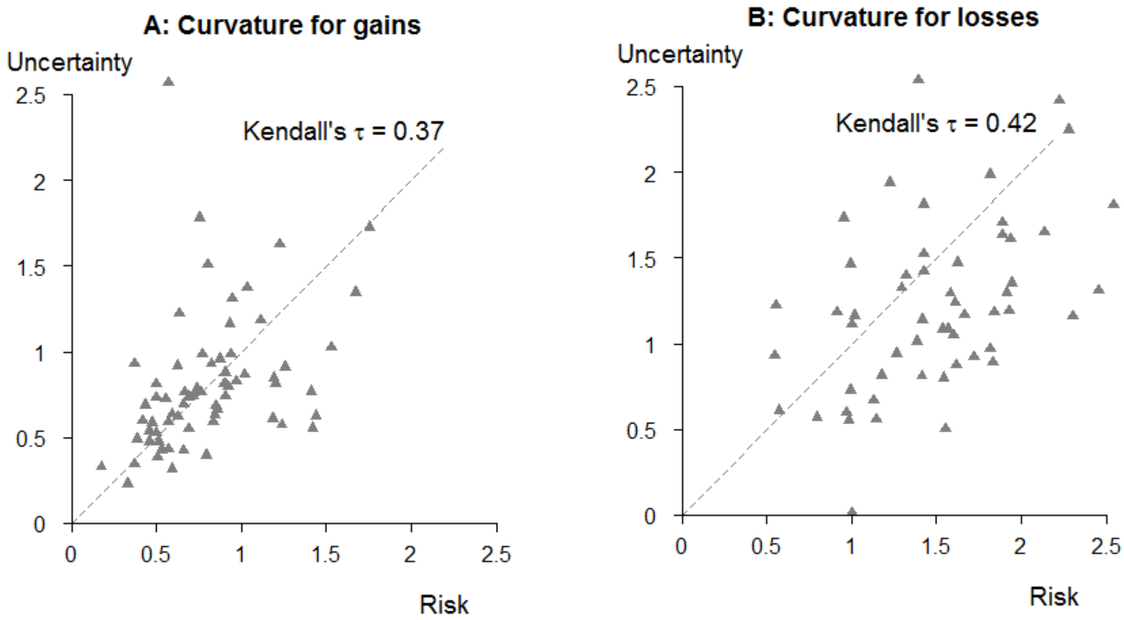


Figure 6.2: Individual Power Coefficients under Risk and Uncertainty

The figure depicts the relationship between individual power coefficients under risk and uncertainty. Panel A displays the power coefficients for gains. Panel B displays the power coefficients for losses. Subjects who had a power coefficient in excess of 2.5 are not shown in the graphs (4 for gains, 7 for losses). The dashed lines correspond to the case where subjects had exactly the same coefficients under risk and uncertainty.

Loss Aversion

Figure 6.3 displays the relationships between the medians of x_j^+ and $-x_j^-$ under risk and under uncertainty. An advantage of our method is that it immediately reveals that there is loss aversion in the sense of Kahneman and Tversky (1979) when $x_j^+ > -x_j^-$.[50] Hence, there is no need to measure the entire utility function to obtain insight into the presence or absence of loss aversion. As Figure 6.3 clearly shows, $-x_j^-$ is below x_j^+ for all j, both under risk and under uncertainty. An estimate of the degree of loss aversion is obtained

[50] For a given j, x_j^+ and $-x_j^-$ have the same utility by construction, $U(x_j^+) = -U(x_j^-)$, and, thus, $x_j^+ > -x_j^-$ implies that $U(x_j^+) < -U(-x_j^+)$, consistent with Kahneman and Tversky's definition of loss aversion ($U(x) < -U(-x)$ for all x > 0).

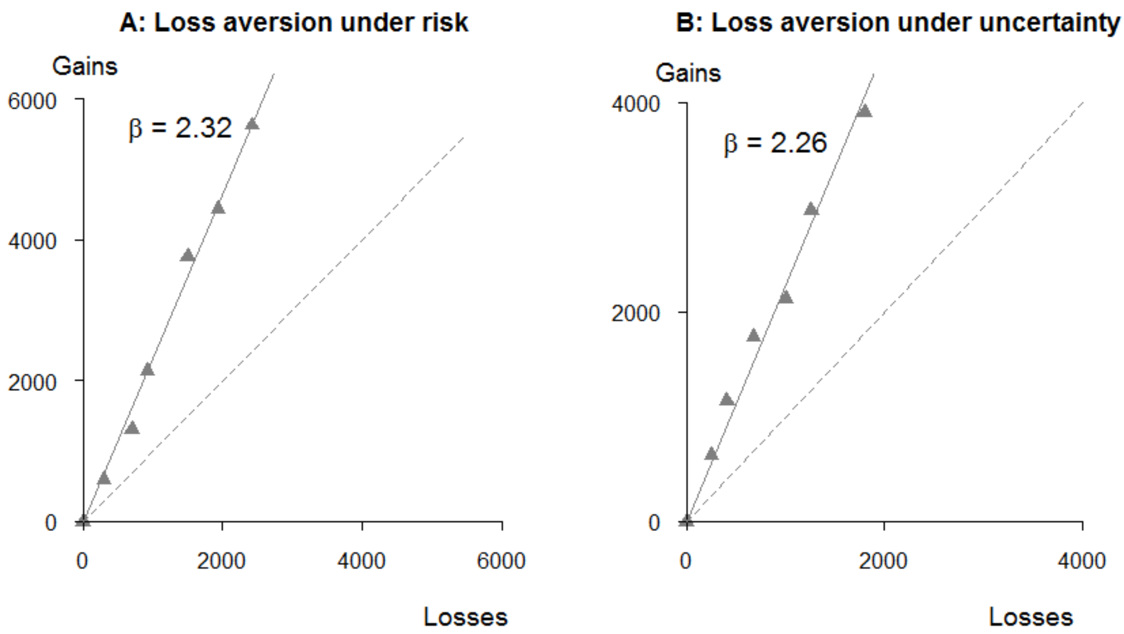


Figure 6.3: Relationship Between Gains and Losses of Same Utility

The figure depicts the relationship between median gains and median losses with the same absolute utility. Panel A displays the relationship between median gains and losses under risk. Panel B displays the same relationship under uncertainty. The dashed line corresponds to the case where gains and losses of the same absolute utility would be of equal size. The straight line with slope corresponds to the best fitting linear equation through all points.

by regressing the x_j^+ on $\left(-x_j^-\right)$. The $\beta's$ in Figure 6.3 display the coefficients from this regression. Both $\beta's$ (for risk and uncertainty) are different from unity ($p < 0.001$) and the values that we obtain are close to those observed previously for risk. We cannot reject the hypothesis that the values of β are the same for risk and uncertainty ($p = 0.431$), which can be taken as an indication that loss aversion is similar under risk and uncertainty.

Moving to the individual level, we find that $x_j^+ > -x_j^-$ for all j (Wilcoxon tests, all $p < 0.001$). Furthermore, $x_j^+/-x_j^-$ does not differ between risk and uncertainty for any j (Mann-Whitney U tests: all $p > 0.254$).

Table 6.4: Results Under the Various Definitions of Loss Aversion

The table depicts the results under the two definitions of loss aversion for both risk and uncertainty. The table displays how the coefficient is defined and the number of loss averse, gain seeking, and loss neutral subjects in both conditions. The numbers in parentheses for Kahneman and Tversky's definition correspond with the case where response errors are not taken into account. Furthermore, the table depicts the median and interquartile range (IQR) for each measure of loss aversion under both definitions.

Definition	Coefficient	Condition	Median [IQR]	Loss averse	Gain seeking	Loss neutral
Kahneman and Tversky (1979)	$\frac{-U(-x)}{U(x)}$	Risk	2.19 [1.06, 5.59]	58(46)	10(6)	1(1)
		Uncertainty	2.48 [1.10, 7.16]	54(50)	16(10)	0(0)
Köbberling and Wakker (2005)	$\frac{x_1^+}{-x_1^-}$	Risk	1.86 [1.06, 4.47]	56	13	3
		Uncertainty	2.00 [1.21, 6.50]	57	14	1

Table 6.4 shows the results of the individual analyses of loss aversion based on Kahneman and Tversky's (1979) and Köbberling and Wakker's (2005) measures. The table clearly shows evidence of loss aversion, irrespective of the definition used and regardless of whether we take response errors into account. According to both definitions, the median loss aversion coefficients for risk and uncertainty do not differ (Mann-Whitney *U* tests: $p > 0.257$ in both tests) and are moderately correlated (Kendall's $\tau > 0.368$, $p < 0.001$ in both tests).

Finally, the two measures of loss aversion are substantially correlated. For risk, Kendall's τ is 0.740 and for uncertainty it is 0.799 (all $p < 0.001$ in both cases). It is comforting to observe that these two distinct measures, one of a local nature and relying on a single kink in the slope of the utility function, and the other global and relying on different absolute utilities associated with the same absolute money amounts in the positive and negative domain, show a high degree of consistency in classifying subjects.

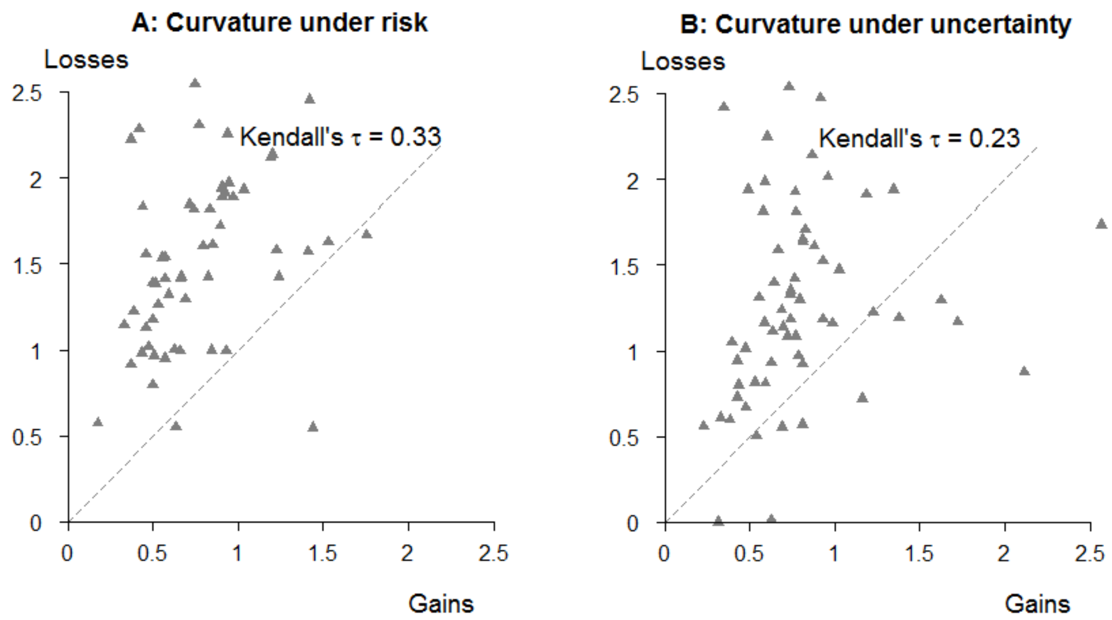


Figure 6.4: The Relationship Between Individual Power Coefficients for Gains and Losses

Panel A displays the power coefficients under risk. Panel B displays the power coefficients under uncertainty. Subjects who had a power coefficient in excess of 2.5 are not shown in the graphs (6 for risk, 9 for uncertainty). The dashed lines correspond to the case where subjects had exactly the same coefficients for gains and losses.

Reflection

The aggregate findings reported earlier suggest that the power coefficients are similar in the gain and loss domains. This implies that the utility for losses is the mirror image of the utility for gains and is referred to as *reflection*.[51] It is of interest to test whether reflection also holds at the individual level. Practically, this would allow us to infer utility for both gains and losses by only measuring it in one of these domains. Theoretically, it would provide support for the idea that utility in both domains is caused by the same psychophysical response to changes relative to the reference point. Reflection is a central

[51]Reflection is also defined as risk [ambiguity] attitudes for losses being the mirror image of risk [ambiguity] attitudes for gains. As risk [ambiguity] attitudes are jointly determined by utility and event weighting under binary PT, it is clear that this definition differs from the one we use here.

result in Tversky and Kahneman (1992) and is widely adopted in theoretical and empirical analyses based on prospect theory (*e.g.,* Barberis, Huang, and Santos, 2001).

We find little indication that reflection should be rejected. Based on the area measure, there was some, albeit marginal, difference in curvature between gains and losses (Mann-Whitney U tests: $p = 0.067$). For uncertainty, there is no difference (Mann-Whitney U tests: $p = 0.724$). Reflection also implies that the power coefficients for gains and losses should be identical. We cannot reject this hypothesis, neither for risk (Mann-Whitney U tests: $p = 0.128$) nor for uncertainty ($p = 0.814$).

On the other hand, both the area measure and the power coefficients, are only slightly correlated under uncertainty, and moderately correlated under risk. For the area measure, Kendall's τ is 0.317 under risk ($p < 0.001$), and 0.191 under uncertainty ($p = 0.018$). For the power coefficients, Kendall's τ is 0.325 under risk ($p < 0.001$), and 0.231 under uncertainty ($p = 0.004$). Figure 6.4 displays the relation between the power coefficients for both risk and uncertainty. The straight line corresponds to reflection. Both for risk and for uncertainty, reflection approximately held for most subjects, but for some it is a poor working hypothesis, particularly under uncertainty.

6.7 Conclusions and Discussion

Ambiguity models differ in whether they allow different utility functions for risk and uncertainty. Under binary prospect theory, which includes the multiple priors models and prospect theory as special cases, utility is independent of the source of uncertainty and, hence, the same for risk and uncertainty. Ambiguity aversion is modelled through a difference in event weighting. We test empirically whether the assumption of identical utility functions is justified and obtain support for it. We cannot reject the hypothesis that utility and loss aversion are the same under risk and under uncertainty. We also obtain convincing evidence for reference-dependence: utility is concave for gains, but convex for losses and there is substantial loss aversion. Finally, the elicited standard sequences are similar for different stimuli supporting the central condition underlying binary prospect theory (Köbberling and Wakker 2003), which has not been tested before.

Our findings pose a descriptive challenge for models that explain ambiguity aversion through a difference in utility curvature between risk and uncertainty alone, like the popular smooth ambiguity model. We observed that standard sequences are similar for risk and uncertainty. In Appendix C we show that this implies under the smooth model that the utility function under uncertainty cannot be a concave or convex transformation of the utility function under risk, even on small preference intervals. Hence, the transformation function has an irregular shape, which complicates its use in applications.

It is interesting that loss aversion under risk and under uncertainty is similar. If loss aversion reflects the psychological intuition that losses loom larger than gains then one would expect that measurements of loss aversion are related across domains. Previous evidence of this correlation gave mixed results. Gächter, Johnson, and Herrmann (2010) found a positive correlation between loss aversion in a risky and in a riskless task, but Abdellaoui *et al.* (2013a) found that loss aversion under risk and loss aversion in intertemporal choice were uncorrelated. Several studies have found that loss aversion is fickle and subject to framing (*e.g.,* Novemsky and Kahneman 2005, Ert and Erev 2008, Abdellaoui *et al.* 2013a). We find that loss aversion is stable under risk and uncertainty if the elicitation method is held constant.

In many decisions probabilities are unknown. People are often not neutral towards ambiguity and it is often important to take ambiguity attitudes into account. Our study contributes to the application of ambiguity models in empirical studies and decision analysis by providing a new parameter-free method to measure utility and loss aversion under uncertainty that is robust to event weighting and that can easily be implemented. Our method extends the trade-off method by allowing for standard sequences that contain both gains and losses and that go through the reference point. It provides a straightforward way of exploring whether decision makers are loss averse without the need to elicit the entire utility function. As stage 1 of our method shows, three elicitations suffice to measure loss aversion in the sense of Köbberling and Wakker (2005) and with a few more measurements loss aversion in the sense of Kahneman and Tversky (1979) can be verified.

Our conclusion that both utility and loss aversion are the same for risk and for uncertainty is not caused by the fact that subjects faced the same stimuli for risk and uncertainty. A simple heuristic that subjects might have used was to simplify the uncertain decision task by assuming that the probability of their preferred color in the ambiguous urn was ½. Then, the decisions under risk and uncertainty would be the same and our conclusions would naturally follow. Our data does not corroborate this hypothesis. The value of the loss L stated in the first stage of our method is significantly lower under ambiguity (Mann-Whitney *U* tests: $p < 0.001$), consistent with ambiguity aversion. Consequently, the subsequent choices that subjects faced were markedly different for risk and uncertainty. Even though the choices were different, the obtained utilities were similar for risk and for uncertainty.

An easy response strategy in the trade-off method is to let the outcomes of the standard sequence increase by the difference between the gauge outcomes ($\mathcal{L}$ and ℓ in the sequence of gains $\mathcal{G}$ and $\mathcal{g}$ in the sequence of losses). This would bias the results in the direction of linear utility. We checked for this heuristic but found little evidence to support it, even allowing for response error.

The trade-off method is chained in the sense that previous responses are used in the elicitation of subsequent choices. Chaining may lead to error propagation, where errors made in one particular choice affect later choices. We checked for the impact of error propagation using the simulation methods developed by Bleichrodt and Pinto (2000) and Abdellaoui, Vossmann, and Weber (2005) In both simulations, we confirmed the conclusions from those studies that the impact of error propagation was negligible.[52] We also repeated the parametric analysis of utility accounting for serial correlation in the error

[52] Bleichrodt, Cillo, and Diecideu (2010) also concluded that error propagation was negligible in their measurements using the trade-off method.

terms.[53] The estimates were identical to the ones reported in Section 6.5. Hence, we conclude that the chained nature of our measurements had no noticeable impact on the results either.

We used hypothetical outcomes because we wanted to detect utility curvature. For small money amounts little utility curvature is usually observed and the equality of utility for risk and for uncertainty would then automatically follow. A second reason for not using real incentives is that we wanted to include losses. Ambiguity attitudes differ between gains and losses and loss aversion is important in explaining risk and ambiguity attitudes. Because we used substantial losses, we could not implement real incentives: it is impossible to find subjects willing to participate in an experiment in which they can lose substantial amounts of money. Given that all but one of the questions involved losses, we could not play out one of the gain questions for real either. Subjects would know immediately which question would be played out for real. The literature on the importance of real incentives is mixed. Most studies found that for small to modest stakes there was little or no effect of using real instead of hypothetical choices for the kind of tasks that we asked our subjects to perform (Bardsley *et al.*, 2010). Therefore, we concluded that the limited potential advantage of using real incentives did not outweigh the advantages of being able to use larger outcomes and losses.

A potentially more important problem in our analyses is that our design did not allow for obtaining non-parametric measurements of the weighting functions. Therefore, we are unable to test whether the weighting functions differ between risk and ambiguity, and cannot conclusively state that the ambiguity aversion that we observe is caused by differences in event weighting. Furthermore, one might worry that the lack of statistically significant results in our present study is due to decision errors (*i.e.,* noise). In the current design, such errors work against models that model ambiguity aversion through differences in utility between risk and ambiguity, and in favor of models that model ambiguity aversion through differences in event weighting. In this sense, our choice for hypothetical choices

[53] We assumed that the error terms followed an AR(1) process $\epsilon_t = \rho\epsilon_{t-1} + u_t$ with u_t normally distributed with expectation 0 and variance σ^2 and estimated this using generalized least squares.

works in favor of the latter type of model, as real incentives tend to reduce variance (Smith and Walker, 1993; Camerer and Hogarth, 1999). In the future, we plan to supplement our analyses with a second experiment that includes measurements on the weighting function and is conducted with a larger group of subjects to increase statistical power.

While the final verdict will thus have to wait, the fact that the curvature and loss aversion parameters are very similar in an absolute sense does pose a descriptive challenge for models that capture ambiguity attitudes through a difference in utility between risk and uncertainty. In addition, our findings convincingly show that reference-dependence of utility is important both in modeling attitudes towards risk and in modeling attitudes towards ambiguity. In both conditions, utility is S-shaped, concave for gains and convex for losses and we observed clear evidence for loss aversion with most subjects being loss averse.

Appendix 6.A Display of the Experimental Questions

Figure 6.5: Choice Screen Under Uncertainty

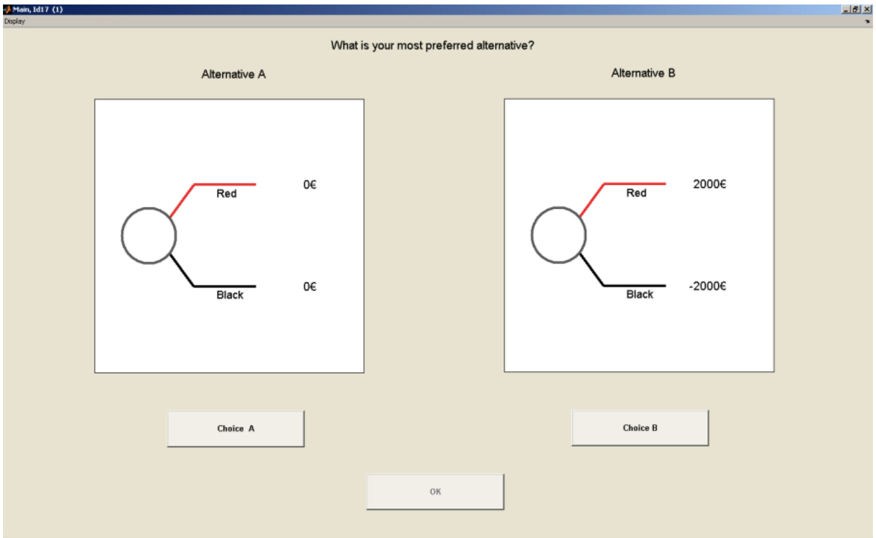


Figure 6.6: Scrollbar Screen Under Uncertainty

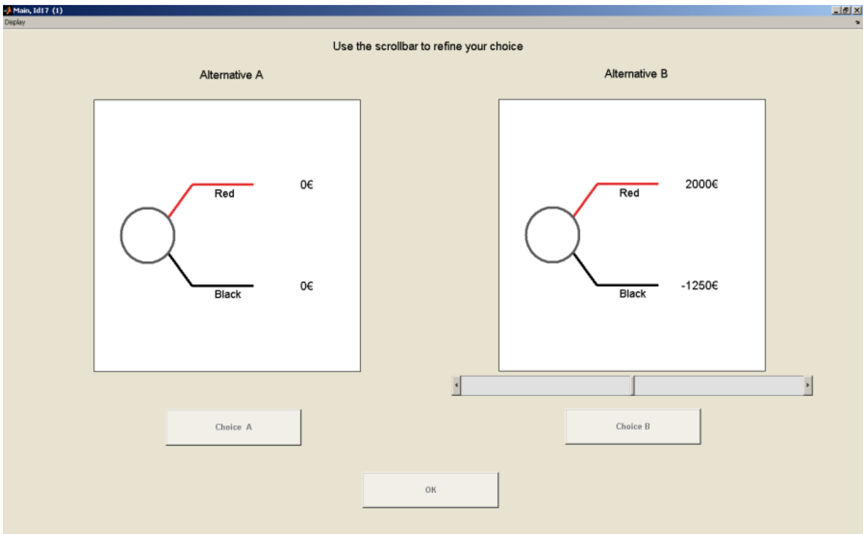
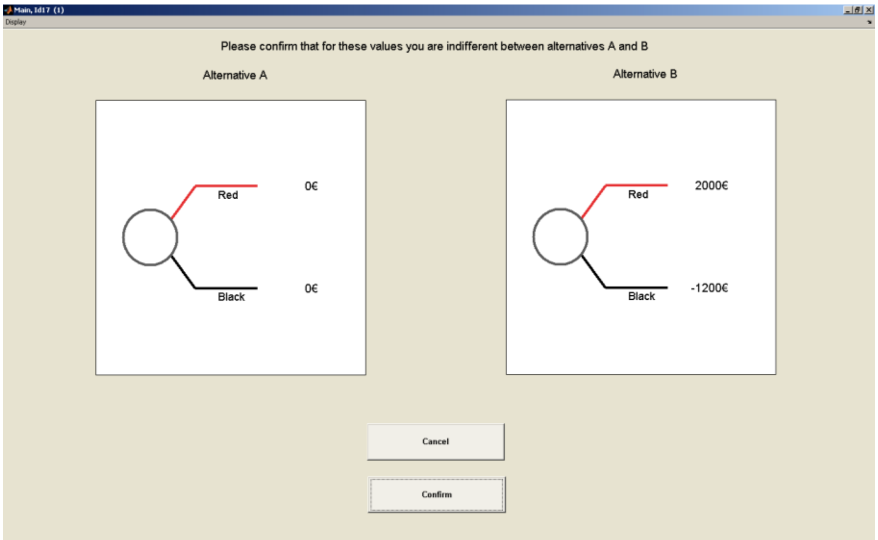


Figure 6.7: Confirmation Screen Under Uncertainty



Appendix 6.B Illustrations of the Bisection Method

Table 6.5: Example Elicitations of Indifferences using the Bisection Method Under Risk

	Offered choices in elicitation L	Offered choices in elicitation x_1^+	Offered choices in elicitation x_2^-
1	**0** vs. (2000, 0.5; -2000)	(2000,0.5;0) vs. **1000**	**(300,0.5;-200)** vs. (800,0.5;-700)
2	0 vs. **(2000, 0.5; -1000)**	**(2000,0.5;0)** vs. 500	**(300,0.5;-200)** vs. (800,0.5;-450)
3	**0** vs. (2000, 0.5; -1500)	(2000,0.5;0) vs. **750**	(300,0.5;-200) vs. **(800,0.5;-325)**
Slider	Start value: -1250 Interval: [-2000,-500]	Start value: 625 Interval: [250,1000]	Start value: -388 Interval: [-576,-200]

Appendix 6.C: Proof Regarding the Smooth Model

The following proof shows that equal utility midpoints for risk and uncertainty imply ambiguity neutrality or volatile ambiguity attitudes under the smooth model. In our experiment we ask indifferences $x_j^+{}_E\mathcal{L} \sim x_{j-1}^+{}_E\ell$. Under the smooth model this implies:

$$\sum_{i=1}^{m} \pi_i \varphi\left(p_i U\left(x_j^+\right) + (1-p_i)U(\mathcal{L})\right) - \sum_{i=1}^{m} \pi_i \varphi\left(p_i U\left(x_{j-1}^+\right) + (1-p_i)U(\ell)\right) =$$

$$\sum_{i=1}^{m} \pi_i \varphi\left(p_i U(x_1^+) + (1-p_i)U(\mathcal{L})\right) - \sum_{i=1}^{m} \pi_i \varphi\left(p_i U(x_0) + (1-p_i)U(\ell)\right) \quad (A1)$$

or:

$$\sum_{i=1}^{m} \pi_i \left(\varphi\left(p_i U\left(x_j^+\right) + (1-p_i)U(\mathcal{L})\right) - \varphi\left(p_i U\left(x_{j-1}^+\right) + (1-p_i)U(\ell)\right)\right) =$$

$$\sum_{i=1}^{m} \pi_i \left(\varphi\left(p_i U(x_1^+) + (1-p_i)U(\mathcal{L})\right) - \varphi\left(p_i U(x_0) + (1-p_i)U(\ell)\right)\right) \quad (A2)$$

Suppose utility midpoints are the same for risk and uncertainty. Because the π_i sum to one, we also have

$$\sum_{i=1}^{m} \pi_i \left(\left(p_i U\left(x_j^+\right) + (1-p_i)U(\mathcal{L})\right) - \left(p_i U\left(x_{j-1}^+\right) + (1-p_i)U(\ell)\right)\right) =$$

$$\sum_{i=1}^{m} \pi_i \left(\left(p_i U(x_1^+) + (1-p_i)U(\mathcal{L})\right) - \left(p_i U(x_0) + (1-p_i)U(\ell)\right)\right) \quad (A3)$$

If φ is strictly concave or strictly convex (A2) and (A3) can never be jointly true. Hence, either φ is linear or it has both convex and concave parts on any interval $[x_{j-1}^+, x_j^+], j = 1, \ldots, k_G$.

Chapter 7 | Conclusions

This thesis employs natural and laboratory experiments to investigate decision making under risk and uncertainty, cooperative behavior, and bargaining.

In Chapter 2 and 3, we find that many of the patterns found in the experimental laboratory carry over to the drastically different environment of TV game shows. In line with previous experimental results, we find evidence that contestants in TV game shows have reciprocal preferences, frame money amounts in relative terms, care about equity, and tend to stick to their promises. The fact that these findings emerge both in low stakes, relatively anonymous laboratory settings with student subjects as well as in high stakes, public game show settings with much more diverse subject pools, is a positive sign with respect to the generalizability of behavioral findings.

At the same time, however, Chapter 4 and 5 show that varying the degree of public scrutiny or presenting subjects with subtle social cues can have a significant impact on behavior. In Chapter 4, we observe that contestants are considerably more risk averse when they make their decisions in the limelight as opposed a more anonymous laboratory setting. In Chapter 5, we observe that presenting subjects with pictures of eyes leads them act in a more social fashion, whereas presenting them with pictures of peers does not uniformly enhance social behavior but does trigger more rational behavior in individual choice tasks.

Comparing these results to those of the first two chapters, this thesis provides a mixed message about the generalizability of findings between different environments. On the one hand, qualitative findings seem to be highly robust across different conditions. In our TV game shows, we observe behavioral patterns that resemble well-documental patterns from the laboratory. Similarly, in Chapter 4 we observe path dependence in risk behavior both in and out of the limelight. On the other hand, however, quantitative estimates appear to be more volatile. These findings are in line with recent studies suggesting that qualitative results generalize between laboratory and field settings even if quantitative results differ

(Kagel and Roth, 2000; Tenorio and Cason, 2002; Healy and Noussair, 2004; Isaac and Schnier, 2005; Antonovics, Arcidiacono and Walsh, 2009; Östling et al., 2011; Bolton, Greiner and Ockenfels, 2013).

Finally, Chapter 6 introduces a new method to measure utility and loss aversion under both risk and uncertainty without the need to introduce simplifying parametric assumptions. Using this method, we are unable to reject the hypotheses that both utility and loss aversion are the same for risk and uncertainty, suggesting that utility primarily reflects attitudes towards outcomes. Both under risk and uncertainty, utility is S-shaped, concave for gains and convex for losses, and there is substantial loss aversion.

Bibliography

Abbink, Klaus, and Abdolkarim Sadrieh. 2009. “The Pleasure of Being Nasty.” *Economics Letters*, 105(3): 306–308.

Abbink, Klaus, and Benedikt Herrmann. 2010. “The Moral Costs of Nastiness.” *Economic Inquiry*, 49(2): 631-633.

Abdellaoui, Mohammed, Aurélien Baillon, Laetitia Placido, and Peter P. Wakker. 2011. “The Rich Domain of Uncertainty: Source Functions and Their Experimental Implementation.” *American Economic Review*, 101(3): 695-723.

Abdellaoui, Mohammed, Frank Vossmann, and Martin Weber. 2005. “Choice-Based Elicitation and Decomposition of Decision Weights for Gains and Losses under Uncertainty.” *Management Science*, 51(9): 1384-1399.

Abdellaoui, Mohammed, Han Bleichrodt, and Corina Paraschiv. 2007. “Measuring Loss Aversion under Prospect Theory: A Parameter-Free Approach.” *Management Science*, 53(10): 1659-1674.

Abdellaoui, Mohammed, Han Bleichrodt, Olivier l'Haridon, and Corina Paraschiv. 2013a. “Is There One Unifying Concept of Utility? An Experimental Comparison of Utility under Risk and Utility over Time.” *Management Science*, 59(9): 2153-2169.

Abdellaoui, Mohammed, Han Bleichrodt, Olivier l’Haridon, and Dennie van Dolder. 2013b. “Source ‐ Dependence of Utility and Loss Aversion: A Critical Test of Ambiguity Models.” Working paper. Available at SSRN: http://ssrn.com/abstract=2318352

Abeler, Johannes, Armin Falk, Lorenz Goette, and David Huffman. 2011. “Reference Points and Effort Provision.” *American Economic Review*, 101(2): 470-492.

Abreu, Dilip, and Faruk Gul. 2000. "Bargaining and Reputation." *Econometrica*, 68(1): 85-117.

Adams, J. Stacy. 1965. "Inequality in Social Exchange." In *Advances in Experimental Social Psychology,* Vol. 2, ed. Leonard Berkowitz, 267-299. Academic Press.

Ai, Chunrong, and Edward C. Norton. 2003. "Interaction Terms in Logit and Probit Models." *Economics Letters*, 80(1): 123-129.

Antonovics, Kate, Peter Arcidiacono, and Randall Walsh. 2005. "Games and Discrimination: Lessons from *The Weakest Link*." *Journal of Human Resources*, 40(4): 918-947.

Antonovics, Kate, Peter Arcidiacono, and Randall Walsh. 2009. "The Effects of Gender Interactions in the Lab and in the Field." *Review of Economics and Statistics*, 91(1): 152-162.

Ariely, Dan, George Loewenstein, and Drazen Prelec. 2003. "Coherent Arbitrariness: Stable Demand Curves Without Stable Preferences." *Quarterly Journal of Economics*, 118(1): 73-105.

Babcock, Linda, and George Loewenstein. 1997. "Explaining Bargaining Impasse: The Role of Self-Serving Biases." *Journal of Economic Perspectives*, 11(1): 109-126.

Babcock, Linda, George Loewenstein, Samuel Issacharoff, and Colin F. Camerer. 1995. "Biased Judgments of Fairness in Bargaining." *American Economic Review*, 85(5): 1337-1343.

Babcock, Linda, Xianghong Wang, and George Loewenstein. 1996. "Choosing the Wrong Pond: Social Comparisons in Negotiations that Reflect a Self-Serving Bias." *Quarterly Journal of Economics*, 111(1): 1-19.

Baillon, Aurélien, Asli Selim, and Dennie van Dolder. 2013. "On the Social Nature of Eyes: The Effect of Social Cues in Interaction and Individual Choice Tasks." *Evolution and Human Behavior*, 34(2): 146-154.

Baltussen, Guido, Martijn J. van den Assem, and Dennie van Dolder. 2013. "Risky Choice in the Limelight." Working paper. Available at SSRN: http://ssrn.com/abstract=2057134

Baltussen, Guido, Thierry Post, Martijn J. van den Assem, and Peter P. Wakker. 2012. "Random Incentive Systems in a Dynamic Choice Experiment." *Experimental Economics*, 15(3): 418-443.

Bar-Hillel, Maya. 1973. "On the Subjective Probability of Compound Events." *Organizational Behavior and Human Performance*, 9(3): 396-409.

Barber, Brad M., and Terrance Odean. 2001. "The Internet and the Investor." *Journal of Economic Perspectives,* 15(1): 41-54.

Barber, Brad M., and Terrance Odean. 2002. "Online Investors: Do the Slow Die First?" *Review of Financial Studies*, 15(2): 455-487.

Barberis, Nicholas, Ming Huang, and Tano Santos. 2001. "Prospect theory and asset prices." *Quarterly Journal of Economics*, 116(1): 1-53.

Barclay, Pat, and Martin L. Lalumière. 2006. "Do People Differentially Remember Cheaters?" *Human Nature*, 17(1): 98-113

Bardsley, Nicholas, Robin Cubitt, Graham Loomes, Peter Moffatt, Chris Starmer, and Robert Sugden. 2010. *Experimental Economics: Rethinking the Rules.* Princeton University Press, Princeton and Oxford.

Bateson, Melissa, Daniel Nettle, and Gilbert Roberts. 2006. "Cues of Being Watched Enhance Cooperation in a Real-World Setting." *Biology Letters*, 2(3): 412–414.

Baucells, Manel, Martin Weber, and Frank Welfens. 2011. "Reference-Point Formation and Updating." *Management Science*, 57(3): 506-519.

Bazerman, Max H., Jared R. Curhan, Don A. Moore, and Kathleen L. Valley. 2000. "Negotiation." *Annual Review of Psychology*, 51: 279-314.

Beetsma, Roel M. W. J., and Peter C. Schotman. 2001. "Measuring Risk Attitudes in a Natural Experiment: Data from the Television Game Show Lingo." *Economic Journal*, 111(474): 821-848.

Bellemare, Charles, and Sabine Kröger. 2007. "On Representative Social Capital." *European Economic Review*, 51(1): 183-202.

Belot, Michèle, V. Bhaskar, and Jeroen van de Ven. 2010a. "Promises and Cooperation: Evidence from a TV Game Show." *Journal of Economic Behavior & Organization*, 73(3): 396-405.

Belot, Michèle, V. Bhaskar, and Jeroen van de Ven. 2010b. "Can Observers Predict Trustworthiness?" *Review of Economics and Statistics*, forthcoming.

Benartzi, Shlomo, and Richard H. Thaler. 1995. "Myopic Loss Aversion and the Equity Premium Puzzle." *Quarterly Journal of Economics*, 110(1): 73-92.

Bennett, Randall W., and Kent A. Hickman. 1993. "Rationality and the 'Price is Right'." *Journal of Economic Behavior & Organization*, 21(1): 99-105.

Berk, Jonathan B., Eric Hughson, and Kirk Vandezande. 1996. "The Price Is Right, But Are the Bids? An Investigation of Rational Decision Theory." *American Economic Review*, 86(4): 954-970.

Binmore, Ken. 1999. "Why Experiment in Economics?" *Economic Journal*, 109(453): 16-24.

Birnbaum, Michael H. 2008. "New Paradoxes of Risky Decision Making." *Psychological Review*, 115(2): 463-501.

Birnbaum, Michael H., and Jeffrey P. Bahra. 2007. "Gain-Loss Separability and Coalescing in Risky Decision Making." *Management Science*, 53(6): 1016-1028.

Blavatskyy, Pavlo, and Ganna Pogrebna. 2010. "Models of Stochastic Choice and Decision Theories: Why Both Are Important for Analyzing Decisions." *Journal of Applied Econometrics*, 25(6): 963-986.

Bleichrodt, Han, Alessandra Cillo, and Enrico Diecidue. 2010. "A Quantitative Measurement of Regret Theory." *Management Science*, 56(1): 161-175.

Bleichrodt, Han, and Jose Luis Pinto. 2000. "A Parameter-Free Elicitation of the Probability Weighting Function in Medical Decision Analysis." *Management Science*, 46(11): 1485-1496.

Blount, Sally. 1995. "When Social Outcomes Aren't Fair: The Effect of Causal Attributions on Preferences." *Organizational Behavior and Human Decision Processes*, 63(2): 131-144.

Bolton, Gary E., and Axel Ockenfels. 2000. "ERC: A Theory of Equity, Reciprocity, and Competition." *American Economic Review*, 90(1): 166-193.

Bolton, Gary, Ben Greiner, and Axel Ockenfels. 2013. "Engineering Trust: Reciprocity in the Production of Reputation Information." *Management Science*, 59(2): 265-285.

Bond, Charles F., and Linda J. Titus. 1983. "Social Facilitation: A Meta-Analysis of 241 Studies." *Psychological Bulletin,* 94(2): 265-292.

Booij, Adam S., and Gijs van de Kuilen. 2009. "A Parameter-Free Analysis of the Utility of Money for the General Population under Prospect Theory." *Journal of Economic Psychology*, 30(4): 651-666.

Booij, Adam S., Bernard M. S. van Praag, and Gijs van de Kuilen. 2010. "A Parametric Analysis of Prospect Theory's Functionals for the General Population." *Theory and Decision*, 68(1-2): 115-148.

Bostic, Raphael, R. J. Herrnstein, and R. Duncan Luce. 1990. “The Effect on the Preference Reversal Phenomenon of Using Choice Indifferences.” *Journal of Economic Behavior & Organization*, 13(2): 193-212.

Bourrat, Pierrick, Nicolas Baumard, and Ryan McKay. 2011. “Surveillance Cues Enhance Moral Condemnation.” *Evolutionary Psychology*, 9(2): 193-199.

Bowles, Hannah Riley, Linda Babcock, and Kathleen L. McGinn. 2005. “Constraints and Triggers: Situational Mechanics of Gender in Negotiation.” *Journal of Personality and Social Psychology*, 89(6): 951-965.

Bowman, David, Deborah Minehart, and Matthew Rabin. 1999. “Loss Aversion in a Consumption-Savings Model.” *Journal of Economic Behavior & Organization*, 38(2): 155-178.

Brandts, Jordi, and Carles Solà. 2001. “Reference Points and Negative Reciprocity in Simple Sequential Games.” *Games and Economic Behavior*, 36(2): 138-157.

Brandts, Jordi, and Gary Charness. 2003. “Truth or Consequences: An Experiment.” *Management Science*, 49(1): 116-130.

Brooks, Robert D., Robert W. Faff, Daniel Mulino, and Richard Scheelings. 2009a. “Deal or No Deal, That Is the Question: The Impact of Increasing Stakes and Framing Effects on Decision-Making under Risk.” *International Review of Finance*, 9(1-2): 27-50.

Brooks, Robert D., Robert W. Faff, Daniel Mulino, and Richard Scheelings. 2009b. “Does Risk Aversion Vary with Decision-Frame? An Empirical Test Using Recent Game Show Data.” *Review of Behavioral Finance*, 1(1-2): 44-61.

Bruhin, Adrian, Helga Fehr-Duda, and Thomas Epper. 2010. “Risk and Rationality: Uncovering Heterogeneity in Probability Distortion. *Econometrica*, 78(4): 1372-1412.

Burnham, Terence C., and Brian Hare. 2007. "Engineering Human Cooperation. Does Involuntary Neural Activation Increase Public Goods Contributions?" *Human Nature*, 18(2): 88-108.

Camerer, Colin F. 2003. *Behavioral Game Theory: Experiments in Strategic Interaction.* Princeton University Press.

Camerer, Colin F. 2011. "The Promise and Success of Lab-Field Generalizability in Experimental Economics: A Critical Reply to Levitt and List." Forthcoming in *The Methods of Modern Experimental Economics*, ed. Guillaume R. Fréchette and Andrew Schotter. Oxford University Press.

Camerer, Colin F., and George Loewenstein. 1993. "Information, Fairness, and Efficiency in Bargaining." In *Psychological Perspectives on Justice: Theory and Applications,* ed. Barbara A. Mellers and Jonathan Baron, 155-179. Cambridge University Press.

Camerer, Colin F., and Martin Weber. 1992. "Recent Developments in Modeling Preferences: Uncertainty and Ambiguity." *Journal of Risk and Uncertainty,* 5(4): 325-370.

Camerer, Colin F., and Robin Hogarth. 1999. "The Effects of Financial Incentives in Experiments: A Review and Capital-Labor-Production Framework". *Journal of Risk and Uncertainty,* 19(1-3): 7-42.

Carpenter, Jeffrey P., Amrita G. Daniere, and Lois M. Takahashi. 2004. "Cooperation, Trust, and Social Capital in Southeast Asian Urban Slums." *Journal of Economic Behavior & Organization*, 55(4): 533-551.

Carpenter, Jeffrey P., Cristina Connolly, and Caitlin Knowles Myers. 2008. "Altruistic Behavior in a Representative Dictator Experiment." *Experimental Economics*, 11(3): 282-298.

Carpenter, Jeffrey P., Eric Verhoogen, and Stephen Burks. 2005. "The Effect of Stakes in Distribution Experiments." *Economics Letters*, 86(3): 393-398.

Charness, Gary, and Martin Dufwenberg. 2005. "Deception: The Role of Guilt." Working paper. UC Santa Barbara and University of Arizona.

Charness, Gary, and Martin Dufwenberg. 2006. "Promises and Partnership." *Econometrica*, 74(6): 1579–1601.

Charness, Gary, and Matthew Rabin. 2002. "Understanding Social Preferences With Simple Tests." *Quarterly Journal of Economics*, 117(3): 817-869.

Charness, Gary. 2004. "Attribution and Reciprocity in an Experimental Labor Market." *Journal of Labor Economics*, 22(3): 665-688.

Chew, Soo Hong, King King Li, Robin Chark, and Songfa Zhong. 2008. "Source Preference and Ambiguity Aversion: Models and Evidence from Behavioral and Neuroimaging Experiments." In *Advances in Health Economics and Health Services Research Volume 20, Neuroeconomics,* ed. Daniel M. Houser, Kevin McGabe, 179-201. Emerald Insight.

Clark, Jeremy. 1998. "Fairness in Public Good Provision: An Investigation of Preferences for Equality and Proportionality." *Canadian Journal of Economics*, 31(3): 708-729.

Cohen, Michele, Jean-Yves Jaffray, and Tanios Said. 1987. "Experimental Comparison of Individual Behavior under Risk and under Uncertainty for Gains and for Losses." *Organizational Behavior and Human Decision Processes*, 39(1): 1-22.

Compte, Olivier, and Philippe Jehiel. 2002. "On the Role of Outside Options in Bargaining with Obstinate Parties." *Econometrica*, 70(4): 1477-1517.

Cooper, David J., and John H. Kagel. 2009. "Other Regarding Preferences: A Selective Survey of Experimental Results." Working paper.

Cosmides, Leda, and John Tooby. 1989. "Evolutionary Psychology and the Generation of Culture: Part II. A Computational Theory of Social Exchange." *Ethology and Sociobiology*, 10(1-3): 51-97.

Cosmides, Leda, and John Tooby. 1992. "Cognitive Adaptations for Social Exchange. In *The adapted mind: evolutionary psychology and the generation of culture*, ed. Jerome H. Barkow, Leda Cosmides, and John Tooby, 163-228. Oxford University Press.

Cosmides, Leda. 1989. "The Logic of Social Exchange: Has Natural Selection Shaped How Humans Reason? Studies with the Watson Selection Task." *Cognition*, 31(3): 187-276.

Cox, James C. 2002. "Trust, Reciprocity, and Other-Regarding Preferences: Groups vs. Individuals and Males vs. Females." In *Experimental Business Research*, ed. Rami Zwick and Amnon Rapoport, 331-350. Kluwer Academic Publishers.

Cox, James C. 2004. "How to Identify Trust and Reciprocity." *Games and Economic Behavior*, 46(2): 260-281.

Cox, James C., Daniel Friedman, and Steven Gjerstad. 2007. "A Tractable Model of Reciprocity and Fairness." *Games and Economic Behavior*, 59(1): 17-45.

Cox, Taylor H., Sharon A. Lobel, and Poppy Lauretta McLeod. 1991. "Effects of Ethnic Group Cultural Differences on Cooperative and Competitive Behavior on a Group Task." *Academy of Management Journal*, 34(4): 827-847.

Crawford, Vincent P. 1982. "A Theory of Disagreement in Bargaining." *Econometrica*, 50(3): 607-637.

Crawford, Vincent P., and Joel Sobel. 1982. "Strategic Information Transmission." *Econometrica*, 50(6): 1431-1451.

Croson, Rachel, and Uri Gneezy. 2009. "Gender Differences in Preferences." *Journal of Economic Literature*, 47(2): 448-474.

Curley, Shawn P., J. Frank Yates, and Richard A. Abrams. 1986. "Psychological Sources of Ambiguity Avoidance." *Organizational Behavior and Human Decision Processes,* 38(2): 230-256.

Dawes, Robyn M. 1980. "Social Dilemmas." *Annual Review of Psychology*, 31: 169-193.

Dawes, Robyn M., and David M. Messick. 2000. "Social Dilemmas." *International Journal of Psychology*, 35(2): 111-116.

Dawes, Robyn M., and Richard H. Thaler. 1988. "Anomalies: Cooperation." *Journal of Economic Perspectives*, 2(3): 187-197.

de Quervain, Dominique J.-F., Urs Fischbacher, Valerie Treyer, Melanie Schellhammer, Ulrich Schnyder, Alfred Buck, and Ernst Fehr. 2004. "The Neural Basis of Altruistic Punishment." *Science*, 305(5688): 1254-1258.

Deck, Cary A., Jungmin Lee, and Javier A. Reyes. 2008. "Risk Attitudes in Large Stake Gambles: Evidence from a Game Show." *Applied Economics*, 40(1): 41-52.

Dijksterhuis, Ap, and Ad van Knippenberg. 1998. The Relation Between Perception and Behavior, or How to Win a Game of Trivial Pursuit. *Journal of Personality and Social Psychology*, 74(4): 865-877.

Du, Ning, and David V. Budescu. 2005. "The Effects of Imprecise Probabilities and Outcomes in Evaluating Investment Options." *Management Science*, 51(12): 1791-1803.

Dufwenberg, Martin, and Astri Muren. 2006. "Generosity, Anonymity, Gender." *Journal of Economic Behavior & Organization*, 61(1): 42-49.

Dufwenberg, Martin, and Georg Kirchsteiger. 2004. "A Theory of Sequential Reciprocity." *Games and Economic Behavior*, 47(2): 268-298.

Dufwenberg, Martin, Simon Gächter, and Heike Hennig-Schmidt. 2010. "The Framing of Games and the Psychology of Play." *Games and Economic Behavior*, forthcoming.

Eckel, Catherine C., and Philip J. Grossman. 2008. "Differences in the Economic Decisions of Men and Women: Experimental Evidence." In *Handbook of Experimental Economics Results; Volume 1*, ed. Charles R. Plott and Vernon L. Smith, 509-519. Elsevier.

Eeckhoudt, Louis, Christian Gollier, and Harris Schlesinger. 1996. “Changes in Background Risk and Risk Taking Behavior.” *Econometrica*, 64(3): 683-689.

Egas, Martijn, and Arno Riedl. 2008. “The Economics of Altruistic Punishment and the Maintenance of Cooperation.” *Proceedings of the Royal Society B: Biological Sciences*, 275(1637): 871-878.

Ekström, Mathias. 2012. “Do Watching Eyes Affect Charitable Giving? Evidence from a Field Experiment.” *Experimental Economics*, 15(3): 530-546.

Elfenbein, Hillary Anger, Jared R. Curhan, Noah Eisenkraft, Aiwa Shirako, and Lucio Baccaro. 2008. “Are Some Negotiators Better Than Others? Individual Differences in Bargaining Outcomes.” *Journal of Research in Personality*, 42(6): 1463-1475.

Ellingsen, Tore, and Topi Miettinen. 2008. “Commitment and Conflict in Bilateral Bargaining.” *American Economic Review*, 98(4): 1629-1635.

Ellsberg, Daniel. 1961. “Risk, Ambiguity, and the Savage Axioms.” *Quarterly Journal of Economics*, 75(4): 643-669.

Engelmann, Dirk, and Guillaume Hollard. 2010. “Reconsidering the Effect of Market Experience on the ‘Endowment Effect’.” *Econometrica*, 78(6): 2005-2019.

Engelmann, Dirk, and Urs Fischbacher. 2009. “Indirect Reciprocity and Strategic Reputation Building in an Experimental Helping Game.” *Games and Economic Behavior*, 67(2): 399–407.

Ericson, Keith M. Marzilli, and Andreas Fuster. 2011. “Expectations as Endowments: Evidence on Reference-Dependent Preferences from Exchange and Valuation Experiments.” *Quarterly Journal of Economics*, 126(4): 1879-1907.

Ernest-Jones, Max, Daniel Nettle, and Melissa Bateson. 2011. “Effects of Eye Images on Everyday Cooperative Behavior: A Field Experiment.” *Evolution and Human Behavior*, 32(3): 172–178.

Ert, Eyal, and Ido Erev. 2008. "The Rejection of Attractive Gambles, Loss Aversion, and the Lemon Avoidance Heuristic." *Journal of Economic Psychology*, 29(5): 715-723.

Falk, Armin, and James J. Heckman. 2009. "Lab Experiments Are a Major Source of Knowledge in the Social Sciences." *Science*, 326(5952): 535-538.

Falk, Armin, and Urs Fischbacher. 2006. "A Theory of Reciprocity." *Games and Economic Behavior*, 54(2): 293-315.

Falk, Armin, Ernst Fehr, and Urs Fischbacher. 2003. "On the Nature of Fair Behavior." *Economic Inquiry*, 41(1): 20-26.

Falk, Armin, Ernst Fehr, and Urs Fischbacher. 2008. "Testing Theories of Fairness-Intentions Matter." *Games and Economic Behavior*, 62(1): 287-303.

Fehr, Ernst, and Frédéric Schneider. 2010. "Eyes are on Us, But Nobody Cares: Are Eye Cues Relevant for Strong Reciprocity?" *Proceedings of The Royal Society B: Biological Sciences*, 277(1686): 1315-1323.

Fehr, Ernst, and Herbert Gintis. 2007. "Human Motivation and Social Cooperation: Experimental and Analytical Foundations." *Annual Review of Sociology*, 33: 43-64.

Fehr, Ernst, and Klaus M. Schmidt. 1999. "A Theory of Fairness, Competition, and Cooperation." *Quarterly Journal of Economics*, 114(3): 817-868.

Fehr, Ernst, and Simon Gächter. 1998. "Reciprocity and Economics: The Economic Implications of Homo Reciprocans." *European Economic Review*, 42(3-5): 845-859.

Fehr, Ernst, and Simon Gächter. 2000a. "Cooperation and Punishment in Public Goods Experiments." *American Economic Review*, 90(4): 980-994.

Fehr, Ernst, and Simon Gächter. 2000b. "Fairness and Retaliation: The Economics of Reciprocity." *Journal of Economic Perspectives*, 14(3): 159-181.

Fehr, Ernst, Urs Fischbacher, and Elena Tougareva. 2013. “Do High Stakes and Competition Undermine Fairness? Evidence from Russia.” Forthcoming in: *Journal of Economic Behavior and Organisation.*

Fischbacher, Urs, and Simon Gächter. 2010. “Social Preferences, Beliefs, and the Dynamics of Free Riding in Public Goods Experiments.” *American Economic Review*, 100(1): 541-556.

Fischbacher, Urs, Simon Gächter, and Ernst Fehr. 2001. “Are People Conditionally Cooperative? Evidence from a Public Goods Experiment.” *Economics Letters*, 71(3): 397-404.

Fox, Craig E., and Amos Tversky. 1995. “Ambiguity Aversion and Comparative Ignorance.” *Quarterly Journal of Economics*, 110(3): 585-603.

Frey, Bruno S., and Stephan Meier. 2004. “Social Comparisons and Pro-Social Behavior: Testing ‘Conditional Cooperation’ in a Field Experiment.” *American Economic Review*, 94(5): 1717-1722.

Friedman, Milton. 1953. *Essays in Positive Economics*. University of Chicago Press, Chicago IL.

Fullenkamp, Connel, Rafael Tenorio, and Robert Battalio. 2003. “Assessing Individual Risk Attitudes Using Field Data from Lottery Games.” *Review of Economics and Statistics*, 85(1): 218-226.

Gächter, Simon, and Arno Riedl. 2005 “Moral Property Rights in Bargaining with Infeasible Claims.” *Management Science*, 51(2): 249-263.

Gächter, Simon, and Benedikt Herrmann. 2009. “Reciprocity, Culture and Human Cooperation: Previous Insights and a New Cross-Cultural Experiment.” *Philosophical Transactions of The Royal Society B: Biological Sciences*, 364(1518): 791-806.

Gächter, Simon, and Ernst Fehr. 1999. “Collective Action as a Social Exchange.” *Journal of Economic Behavior and Organization*, 39(4): 341–369.

Gächter, Simon, Benedikt Herrmann, and Christian Thöni. 2004. “Trust, Voluntary Cooperation, and Socio-Economic Background: Survey and Experimental Evidence.” *Journal of Economic Behavior & Organization*, 55(4): 505-531.

Gächter, Simon, Benedikt Herrmann, and Christian Thöni. 2010. “Culture and Cooperation.” *Philosophical Transactions of The Royal Society B: Biological Sciences*, 365(1553): 2651-2661.

Gächter, Simon, Benedikt Herrmann. 2011. “The Limits of Self-Governance When Cooperators Get Punished: Experimental Evidence from Urban and Rural Russia.” *European Economic Review*, 55(2): 193-210.

Gächter, Simon, Eric. J. Johnson, and Andreas Herrmann. 2010. “Individual-Level Loss Aversion in Risky and Riskless Choice. CeDEx discussion paper series 2010-20

Gajdos, Thibault, Takashi Hayashi, Jean-Marc Tallon, and Jean-Christophe Vergnaud. 2008. “Attitude Toward Imprecise Information.” *Journal of Economic Theory*, 140(1): 27-65.

Gardner, Andy, and Stuart A. West. 2004. “Cooperation and Punishment, Especially in Humans.” *American Naturalist*, 164(6): 753-764.

Gertner, Robert. 1993. “Game Shows and Economic Behavior: Risk-Taking on ‘Card Sharks’.” *Quarterly Journal of Economics*, 108(2): 507-521.

Ghirardato, Paolo, and Massimo Marinacci. 2001. “Risk, Ambiguity, and the Separation of Utility and Beliefs.” *Mathematics of Operations Research*, 26(4): 864-890.

Ghirardato, Paolo, Fabio Maccheroni, and Massimo Marinacci. 2004. “Differentiating Ambiguity and Ambiguity Attitude.” *Journal of Economic Theory*, 118(2): 133-173.

Gilboa, Itzhak, and David Schmeidler. 1989. "Maxmin Expected Utility with Non-Unique Prior." *Journal of Mathematical Economics*, 18(2): 141-153.

Gilboa, Itzhak. 1987. "Expected Utility with Purely Subjective Non-Additive Probabilities." *Journal of Mathematical Economics*, 16(1): 65-88.

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

Gollier, Christian, and John W. Pratt. 1996. "Risk Vulnerability and the Tempering Effect of Background Risk." *Econometrica*, 64(5): 1109-1123.

Green, Donald, Karen E. Jacowitz, Daniel Kahneman, and Daniel McFadden. 1998. "Referendum Contingent Valuation, Anchoring, and Willingness to Pay for Public Goods." *Resource and Energy Economics*, 20(2): 85-116.

Gul, Faruk, and Wolfgang Pesendorfer. 2010. "Interdependent Preference Models as a Theory of Intentions." Working paper.

Haley, Kevin J., and Daniel M. T. Fessler. 2005. "Nobody's Watching? Subtle Cues Affect Generosity in an Anonymous Economic Game." *Evolution and Human Behavior*, 26(3): 245-256.

Harrison, Glenn W., and John A. List. 2004. "Field Experiments." *Journal of Economic Literature*, 42(4): 1009-1055.

Harrison, Glenn W., John A. List, and Charles Towe. 2007. "Naturally Occurring Preferences and Exogenous Laboratory Experiments: A Case Study of Risk Aversion." *Econometrica*, 75(2): 433-458.

Hartley, Roger, Gauthier Lanot, and Ian Walker. 2006. *Who Really Wants to Be a Millionaire? Estimates of Risk Aversion from Gameshow Data.* Keele Economics Research Papers 2006/07, Centre for Economic Research, Keele University.

Healy, Paul J., and Charles N. Noussair. 2004. "Bidding Behavior in the *Price Is Right* Game: An Experimental Study." *Journal of Economic Behavior & Organization*, 54(2): 231-247.

Heath, Chip, and Amos Tversky. 1991. "Preference and Belief: Ambiguity and Competence in Choice under Uncertainty." *Journal of Risk and Uncertainty*, 4(1): 5-28.

Henrich, Joseph, Robert Boyd, Samuel Bowles, Colin F. Camerer, Ernst Fehr, and Herbert Gintis (ed.). 2004. *Foundations of Human Sociality: Economic Experiments and Ethnographic Evidence from Fifteen Small-Scale Societies.* Oxford University Press.

Henrich, Joseph, Robert Boyd, Samuel Bowles, Colin F. Camerer, Ernst Fehr, Herbert Gintis, and Richard McElreath. 2001. "In Search of Homo Economicus: Behavioral Experiments in 15 Small-Scale Societies." *American Economic Review*, 91(2): 73-78.

Henrich, Joseph, Steven J. Heine, and Ara Norenzayan. 2010. "The Weirdest People in the World?" *Behavioral and Brain Sciences*, 33(2-3): 61-83.

Herberich, David H., and John A. List. 2012. "Digging into Background Risk: Experiments with Farmers and Students." *American Journal of Agricultural Economics*, 94(2): 457-463.

Herrmann, Benedikt, Christian Thöni, and Simon Gächter. 2008. "Antisocial Punishment Across Societies." *Science*, 319(5868): 1362-1367.

Hersch, Philip L., and Gerald S. McDougall. 1997. "Decision Making under Uncertainty When the Stakes Are High: Evidence from a Lottery Game Show." *Southern Economic Journal*, 64(1): 75-84.

Hoffman, Elizabeth, Kevin A. McCabe, and Vernon L. Smith. 1996a. "On Expectations and the Monetary Stakes in Ultimatum Games." *International Journal of Game Theory*, 25(3): 289-301.

Hoffman, Elizabeth, Kevin A. McCabe, and Vernon L. Smith. 1996b. "Social Distance and Other-Regarding Behavior in Dictator Games." *American Economic Review*, 86(3): 653-660.

Hogarth, Robin M., and Howard Kunreuther. 1989. "Risk, Ambiguity, and Insurance." *Journal of Risk and Uncertainty*, 2(1): 5-35.

Homans, George C. 1958. "Social Behavior as Exchange." *American Journal of Sociology*, 63(6): 597-606.

Huber, Joel, John W. Payne, and Christopher Puto. 1982. "Adding Asymmetrically Dominated Alternatives: Violations of Regularity and the Similarity Hypothesis." *Journal of Consumer Research*, 9(1): 90-98.

Isaac, R. Mark, and Kurt Schnier. 2005. "Silent Auctions in the Field and in the Laboratory." *Economic Inquiry*, 43(4): 715-733.

Jaeggi, Adrian V., Judith M. Burkhart, and Carel P. Van Schaik. 2010. "On the Psychology of Cooperation in Humans and other Primates: Combining the Natural History and Experimental Evidence of Prosociality." *Philosophical Transactions of The Royal Society B: Biological Sciences*, 365(1553): 2723-2735.

Jaffray, Jean-Yves. 1989. "Linear Utility Theory for Belief Functions." *Operations Research Letters*, 8(2): 107-112.

Johansson-Stenman, Olof, Minhaj Mahmud, and Peter Martinsson. 2005. "Does Stake Size Matter in Trust Games?" *Economics Letters*, 88(3): 365-369.

Johnson, Eric J., and David A. Schkade. 1989. "Bias in Utility Assessments: Further Evidence and Explanations." *Management Science*, 35(4): 406-424.

Johnson, Eric J., Gerald Häubl, and Anat Keinan. 2007. "Aspects of Endowment: A Query Theory of Value Construction." *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 33(3): 461-474.

Kagel, John H., and Alvin E. Roth. 2000. “The Dynamics of Reorganization in Matching Markets: A Laboratory Experiment Motivated by a Natural Experiment.” *Quarterly Journal of Economics*, 115(1): 201-235.

Kagel, John H., and Katherine Willey Wolfe. 2001. “Tests of Fairness Models Based on Equity Considerations in a Three-Person Ultimatum Game.” *Experimental Economics*, 4(3): 203-219.

Kahneman, Daniel, and Amos Tversky. 1979. “Prospect Theory: An Analysis of Decision under Risk.” *Econometrica*, 47(2): 263-292.

Kahneman, Daniel, Ilana Ritov, and David A. Schkade. 1999. “Economic Preferences or Attitude Expressions?: An Analysis of Dollar Responses to Public Issues.” *Journal of Risk and Uncertainty*, 19(1-3): 203-235.

Kahneman, Daniel, Jack L. Knetsch, and Richard H. Thaler. 1986. “Fairness and the Assumptions of Economics.” *Journal of Business*, 59(4): 285-300.

Kambe, Shinsuke. 1999. “Bargaining with Imperfect Commitment.” *Games and Economic Behavior*, 28(2): 217-237.

Kameda, Tatsuya, and James H. Davis. 1990. “The Function of the Reference Point in Individual and Group Risk Decision Making.” *Organizational Behavior and Human Decision Processes*, 46(1): 55-76.

Kerr, Norbert L. 1999. “Anonymity and Social Control in Social Dilemmas.” In *Resolving Social Dilemmas*, ed. Margaret Foddy, Michael Smithson, Sherry Schneider and Michael Hogg, 103-119, Psychology Press.

Kessler, Judd B., and Lise Vesterlund. 2012. The External Validity of Laboratory Experiments: The Misleading Emphasis on Quantitative Effects. Forthcoming in *The Methods of Modern Experimental Economics*, ed. Guillaume R. Fréchette and Andrew Schotter, Oxford University Press.

Kiyonari, Toko, Shigehito Tanida, and Toshio Yamagishi. 2000. "Social Exchange and Reciprocity: Confusion or a Heuristic?" *Evolution and Human Behavior*, 21(6): 411–427.

Klibanoff, Peter, Massimo Marinacci, and Sujoy Mukerji. 2005. "A Smooth Model of Decision Making under Ambiguity." *Econometrica*, 73(6): 1849-1892.

Köbberling, Veronika, and Peter P. Wakker. 2003. "Preference Foundations for Nonexpected Utility: A Generalized and Simplified Technique." *Mathematics of Operations Research*, 28(3): 395-423.

Köbberling, Veronika, and Peter P. Wakker. 2005. "An Index of Loss Aversion." *Journal of Economic Theory*, 122(1): 119-131.

Kocher, Martin G., Peter Martinsson, and Martine Visser. 2008. "Does Stake Size Matter for Cooperation and Punishment?" *Economics Letters*, 99(3): 508-511.

Konana, Prabhudev, and Sridhar Balasubramanian. 2005. "The Social-Economic-Psychological Model of Technology Adoption and Usage: An Application to Online Investing." *Decision Support Systems,* 39(3): 505-524.

Konow, James. 1996. "A Positive Theory of Economic Fairness." *Journal of Economic Behavior & Organization*, 31(1): 13-35.

Konow, James. 2000. "Fair Shares: Accountability and Cognitive Dissonance in Allocation Decisions." *American Economic Review*, 90(4): 1072-1091.

Konow, James. 2003. "Which Is the Fairest One of All? A Positive Analysis of Justice Theories." *Journal of Economic Literature*, 41(4): 1188-1239.

Kőszegi, Botond, and Matthew Rabin. 2006. "A Model of Reference-Dependent Preferences." *Quarterly Journal of Economics*, 121(4): 1133-1165.

Kőszegi, Botond, and Matthew Rabin. 2007. "Reference-Dependent Risk Attitudes." *American Economic Review*, 97(4): 1047-1073.

Kruglanski, Arie W., and Tallie Freund. 1983. "The Freezing and Unfreezing of Lay-Inferences: Effects on Impressional Primacy, Ethnic Stereotyping, and Numerical Anchoring." *Journal of Experimental Social Psychology*, 19(5): 448-468.

Kuhnen, Camelia M., and Brian Knutson. 2011. "The Influence of Affect on Beliefs, Preferences, and Financial Decisions." *Journal of Financial and Quantitative Analysis*, 46(3): 605-626.

Lamba, Shakti, and Ruth Mace. 2010. "People Recognise When They Are Really Anonymous in an Economic Game." *Evolution and Human Behavior*, 31(4): 271–278.

Ledyard, John O. 1995. "Public Goods: A Survey of Experimental Research." In *Handbook of Experimental Economics*, ed. John H. Kagel and Alvin E. Roth, 111-194. Princeton University Press.

Lee, Jinkwon. 2008. "The Effect of the Background Risk in a Simple Chance Improving Decision Model." *Journal of Risk and Uncertainty*, 36(1): 19-41.

Lerner, Jennifer S., and Philip E. Tetlock. 1999. "Accounting for the Effects of Accountability." *Psychological Bulletin,* 125(2): 255-275.

Levine, David K. 1998. "Modeling Altruism and Spitefulness in Experiments." *Review of Economic Dynamics*, 1(3): 593-622.

Levitt, Steven D. 2004. "Testing Theories of Discrimination: Evidence from Weakest Link." *Journal of Law and Economics*, 47(2): 431-453.

Levitt, Steven D., and John A. List. 2007a. "What Do Laboratory Experiments Measuring Social Preferences Reveal About the Real World?" *Journal of Economic Perspectives*, 21(2): 153-174.

Levitt, Steven D., **and John A. List.** 2007b. "Viewpoint: On the Generalizability of Lab Behaviour to the Field." *Canadian Journal of Economics*, 40(2): 347-370.

Levitt, Steven D., and John A. List. 2008. “Homo Economicus Evolves.” *Science*, 319(5865): 909-910.

Lipsey, Richard G. 1979. *An Introduction to Positive Economics, 5th edition.* Weidenfeld and Nicholson, London UK.

List, John A. 2003. “Does Market Experience Eliminate Market Anomalies?” *Quarterly Journal of Economics*, 118(1): 41-71.

List, John A. 2004a. “Young, Selfish and Male: Field Evidence of Social Preferences.” *Economic Journal*, 114(492): 121-149.

List, John A. 2004b. “Neoclassical Theory Versus Prospect Theory: Evidence from the Marketplace.” *Econometrica*, 72(2): 615-625.

List, John A. 2006. “Friend or Foe? A Natural Experiment of the Prisoner’s Dilemma.” *Review of Economics and Statistics*, 88(3): 463-471.

List, John A. 2011. “Does Market Experience Eliminate Market Anomalies? The Case of Exogenous Market Experience.” *American Economic Review*, 101(3): 313-317.

List, John A., and Todd L. Cherry. 2000. “Learning to Accept in Ultimatum Games: Evidence from an Experimental Design that Generates Low Offers.” *Experimental Economics*, 3(1): 11-29.

List, John A., and Todd L. Cherry. 2008. “Examining the Role of Fairness in High Stakes Allocation Decisions.” *Journal of Economic Behavior & Organization*, 65(1): 1-8.

Loewenstein, George F. 1999. “Experimental Economics from the Vantage-Point of Behavioural Economics.” *Economic Journal*, 109(453): F25-F34.

Loewenstein, George F., Elke U. Weber, Christopher K. Hsee, and Ned Welch. 2001. “Risk as Feelings.” *Psychological Bulletin,* 127(2): 267-286.

Loewenstein, George, Samuel Issacharoff, Colin F. Camerer, and Linda Babcock. 1993. "Self-Serving Assessments of Fairness and Pretrial Bargaining." *Journal of Legal Studies,* 22(1): 135-159.

Luce, R. Duncan. 1991. "Rank- and Sign-Dependent Linear Utility Models for Binary Gambles." *Journal of Economic Theory*, 53(1): 75-100.

Lusk, Jayson L., and Keith H. Coble. 2008. "Risk Aversion in the Presence of Background Risk: Evidence from an Economic Experiment." In *Risk Aversion in Experiments* (*Research in Experimental Economics*, Volume 12), ed. James C. Cox and Glenn W. Harrison, Emerald Group Publishing Limited, 315-340.

Maccheroni, Fabio, Massimo Marinacci, and Aldo Rustichini. 2006. "Ambiguity Aversion, Robustness, and the Variational Representation of Preferences." *Econometrica*, 74(6): 1447-1498.

Maner, Jon K., J. Anthony Richey, Kiara Cromer, Mike Mallott, Carl W. Lejuez, Thomas E. Joiner and Norman B. Schmidt. 2007. "Dispositional Anxiety and Risk-Avoidant Decision-Making." *Personality and Individual Differences,* 42(4): 665-675.

Mano, Haim. 1994. "Risk-Taking, Framing Effects, and Affect." *Organizational Behavior and Human Decision Processes*, 57(1): 38-58.

Marwell, Gerald, and Ruth E. Ames. 1979. "Experiments on the Provision of Public Goods. I. Resources, Interest, Group Size, and the Free-Rider Problem." *American Journal of Sociology*, 84(6): 1335-1360.

Marwell, Gerald, and Ruth E. Ames. 1980. "Experiments on the Provision of Public Goods. II. Provision Points, Stakes, Experience, and the Free-Rider Problem." *American Journal of Sociology*, 85(4): 926-937.

McCabe, Kevin A., Mary L. Rigdon, and Vernon L. Smith. 2003. "Positive Reciprocity and Intentions in Trust Games." *Journal of Economic Behavior & Organization*, 52(2): 267-275.

Mealey, Linda, Christopher Daood, and Michael Krage. 1996. “Enhanced Memory for Faces of Cheaters.” *Ethology and Sociobiology*, 17(2): 119-128.

Mehl, Bettina, and Axel Buchner. 2008. “No Enhanced Memory for Faces of Cheaters.” *Evolution and Human Behavior*, 29(1): 35-41

Metrick, Andrew. 1995. “A Natural Experiment in ‘Jeopardy’.” *American Economic Review*, 85(1): 240-253.

Mifune, Nobuhiro, Hirofumi Hashimoto, and Toshio Yamagishi. 2010. “Altruism Toward In-Group Members as a Reputation Mechanism.” *Evolution and Human Behavior*, 31(2): 109-117.

Milinsky, Manfred, Dirk Semmann, and Hans-Jürgen Krambeck. 2002. “Reputation Helps Solve the ‘Tragedy of the Commons’.” *Nature*, 415(6870): 424-426.

Milinsky, Manfred, Dirk Semmann, Theo C.M. Bakker, and Hans-Jürgen Krambeck. (2001). “Cooperation Through Indirect Reciprocity: Image Scoring or Standing Strategy?” *Proceedings of The Royal Society B: Biological Sciences,* 268(1484): 2495-2501.

Miller, Luis, and Christoph Vanberg. 2013. “Decision Costs in Legislative Bargaining: An Experimental Analysis.” *Public Choice*, 155(3-4): 373-394.

Miller, Paul M., and N.S. Fagley. 1991. “The Effects of Framing, Problem Variations, and Providing Rationale on Choice.” *Personality and Social Psychology Bulletin,* 17(5): 517-522.

Miyamoto, John M. 1988. “Generic Utility Theory: Measurement Foundations and Applications in Multiattribute Utility Theory.” *Journal of Mathematical Psychology*, 32(4): 357-404.

Munier, Bertrand, and Costin Zaharia. 2002. “High Stakes and Acceptance Behavior in Ultimatum Bargaining: A Contribution from an International Experiment.” *Theory and Decision*, 53(3): 187- 207.

Muthoo, Abhinay. 1996. “A Bargaining Model Based on the Commitment Tactic.” *Journal of Economic Theory*, 69(1): 134-152.

Muthoo, Abhinay. 1999. *Bargaining Theory with Applications*. Cambridge University Press.

Muthukrishnan, A. V., Luc Wathieu, and Alison Jing Xu. 2009. “Ambiguity Aversion and the Preference for Established Brands.” *Management Science,* 55(12): 1933-1941.

Nau, Robert F. 2006. “Uncertainty Aversion with Second-Order Utilities and Probabilities.” *Management Science*, 52(1): 136-145.

Neilson, William S. 2002. “Comparative Risk Sensitivity with Reference-Dependent Preferences.” *Journal of Risk and Uncertainty*, 24(2): 131-142.

Neilson, William S. 2010. “A Simplified Axiomatic Approach to Ambiguity Aversion.” *Journal of Risk and Uncertainty*, 41(2): 113-124.

Nettle, Daniel, Zoe Harper, Adam Kidson, Rosie Stone, Ian S. Penton-Voak, and Melissa Bateson. 2013. “The Watching Eyes Effect in the Dictator Game: It's Not How Much You Give, It's Being Seen to Give Something.” *Evolution and Human Behavior*, 34(1): 35-40.

Noussair, Charles S., Stephane Robin, and Bernard Ruffieux. 2004. “Revealing Consumers' Willingness-to-Pay: A Comparison of the BDM Mechanism and the Vickrey Auction.” *Journal of Economic Psychology*, 25(6): 725-741.

Novemsky, Nathan, and Daniel Kahneman. 2005. “The Boundaries of Loss Aversion.” *Journal of Marketing Research*, 42(2): 119-128.

Nydegger, Rudy V., and Guillermo Owen. 1975. "Two-Person Bargaining: An Experimental Test of the Nash Axioms." *International Journal of Game Theory,* 3(4): 239-249.

Oberholzer-Gee, Felix, Joel Waldfogel, and Matthew W. White. 2010. "Friend or Foe? Cooperation and Learning in High-Stakes Games." *Review of Economics and Statistics*, 92(1): 179-187.

Ochs, Jack, and Alvin E. Roth. 1989. "Experimental Study of Sequential Bargaining." *American Economic Review*, 79(3): 355-384.

Oda, Ryo, and Shun Nakajima. 2010. "Biased Face Recognition in the Faith Game." *Evolution and Human Behavior*, 31(2): 118-122.

Oda, Ryo, Yuki Niwa, Atsushi Honma, and Kai Hiraishi. 2011. "An Eye-Like Painting Enhances the Expectation of a Good Reputation." *Evolution and Human Behavior*, 32(3): 166-171.

Oda, Ryo. 1997. "Biased Face Recognition in the Prisoner's Dilemma Game." *Evolution and Human Behavior*, 18(5): 309-315.

Offerman, Theo. 2002. "Hurting Hurts More Than Helping Helps." *European Economic Review*, 46(8): 1423-1437.

Office for National Statistics. 2009. Statistical bulletin: 2009 Annual survey of hours and earnings. Office for National Statistics, Newport, UK. http://www.ons.gov.uk/ons/rel/ashe/annual-survey-of-hours-and-earnings/2009-revised/stb-ashe-2009.pdf.

Office for National Statistics. 2010. Statistical bulletin: 2010 Annual survey of hours and earnings. Office for National Statistics, Newport, UK. http://www.ons.gov.uk/ons/rel/ashe/annual-survey-of-hours-and-earnings/2010-results/stb-ashe-2010.pdf.

Östling, Robert, Joseph Tao-yi Wang, Eileen Y. Chou, and Colin F. Camerer. 2011. "Testing Game Theory in the Field: Swedish LUPI Lottery Games." *American Economic Journal: Microeconomics*, 3(3): 1-33.

Ostrom, Elinor, James Walker, and Roy Gardner. 1992. "Covenants With and Without a Sword: Self-Governance is Possible." *American Political Science Review*, 86(2): 404-417.

Pennings, Joost M.E., and Ale Smidts. 2003. "The Shape of Utility Functions and Organizational Behavior." *Management Science*, 49(9): 1251-1263.

Post, Thierry, Martijn J. van den Assem, Guido Baltussen, and Richard H. Thaler. 2008. "Deal or No Deal? Decision Making under Risk in a Large-Payoff Game Show." *American Economic Review*, 98(1): 38-71.

Powell, Kate L., Gilbert Roberts, and Daniel Nettle. 2012. "Eye Images Increase Charitable Donations: Evidence From an Opportunistic Field Experiment in a Supermarket." *Ethology*, 118(11): 1096-1011.

Pratt, John W., and Richard J. Zeckhauser. 1987. "Proper Risk Aversion." *Econometrica*, 55(1): 143-154.

Rabin, Matthew. 1993. "Incorporating Fairness into Game Theory and Economics." *American Economic Review*, 83(5): 1281-1302.

Rabin, Matthew. 2000. "Risk Aversion and Expected-utility Theory: A Calibration Theorem." *Econometrica*, 68(5): 1281-1292.

Raghunathan, Rajagopal, and Michel Tuan Pham. 1999. "All Negative Moods Are Not Equal: Motivational Influences of Anxiety and Sadness on Decision Making." *Organizational Behavior and Human Decision Processes*, 79(1): 56-77.

Raiffa, Howard. 1961. "Risk, Ambiguity, and the Savage Axioms: Comment." *Quarterly Journal of Economics,* 75(4): 690–694.

Raihani, Nichola J., and Redouan Bshary. 2012. "A Positive Effect of Flowers Rather Than Eye Images in a Large-Scale, Cross-Cultural Dictator Game." *Proceedings of The Royal Society B: Biological Sciences*, 279(1742): 3556-3564.

Rapoport, Anatol. 1988. "Experiments With N-Person Social Traps I: Prisoner's Dilemma, Weak Prisoner's Dilemma, Volunteer's Dilemma, and Largest Number." *Journal of Conflict Resolution*, 32(3): 457-472.

Raub, Werner, and Chris Snijders. 1997. "Gains, Losses, and Cooperation in Social Dilemmas and Collective Action: The Effects of Risk Preferences." *Journal of Mathematical Sociology*, 22(3): 263-302.

Rege, Mari, and Kjetil Telle. 2004. "The Impact of Social Approval and Framing on Cooperation in Public Good Situations." *Journal of Public Economics*, 88(7-8): 1625-1644.

Rick, Scott, and George Loewenstein. 2008. "The Role of Emotion in Economic Behavior." In *Handbook of Emotions, 3rd edition,* ed. Michael Lewis, Jeannette M. Haviland-Jones and Lisa Feldman Barrett, Guilford Press, New York, 138-156.

Rigdon, Mary, Keiko Ishii, Motoki Watabe, and Shinobu Kitayama. 2009. "Minimal Social Cues in the Dictator Game." *Journal of Economic Psychology*, 30(3): 358-367.

Roca, Mercè, Robin M. Hogarth, A. John Maule. 2006. "Ambiguity Seeking as a Result of the Status Quo Bias." *Journal of Risk and Uncertainty*, 32(3): 175-194.

Rubin, Jeffrey Z., and Bert R. Brown. 1975. "Bargainers as Individuals." In *The Social Psychology of Bargaining and Negotiation*. New York: Academic Press, 157-196.

Rubinstein, Ariel. 1982. "Perfect Equilibrium in a Bargaining Model." *Econometrica*, 50(1): 97-109.

Sally, David. 1995. "Conversation and Cooperation in Social Dilemmas. A Meta-Analysis of Experiments from 1958 to 1992." *Rationality and Society*, 7(1): 58-92.

Samuelson, Paul, and William Nordhaus. 1985. *Principles of Economics, 12th Edition.* McGraw-Hill, New York NY.

Schelling, Thomas C. 1956. "An Essay on Bargaining." *American Economic Review*, 46(3): 281-306.

Schlicht, Ekkehart. 1998. On Custom in the Economy. Clarendon Press, Oxford, U.K.

Schmeidler, David. 1989. "Subjective Probability and Expected Utility Without Additivity." Econometrica 57(3): 571-587.

Schokkaert, Erik, and Bert Overlaet. 1989. "Moral Intuitions and Economic Models of Distributive Justice." *Social Choice and Welfare*, 6(1), 19-31.

Sears, David O. 1986. "College Sophomores in the Laboratory: Influences of a Narrow Data Base on Social Psychology's View of Human Nature." *Journal of Personality and Social Psychology*, 51(3): 515-530.

Seinen, Ingrid, and Arthur Schram. 2006. "Social Status and Group Norms: Indirect Reciprocity in a Helping Experiment." *European Economic Review*, 50(3): 581-602.

Seo, Kyoungwon. 2009. "Ambiguity and Second-Order Belief." *Econometrica*, 77(5): 1575-1605.

Seru, Amit, Tyler Shumway, and Noah Stoffman. 2010. "Learning by Trading." *Review of Financial Studie*s, 23(2): 705-739.

Siegel-Jacobs, Karen, and J. Frank Yates. 1996. "Effects of Procedural and Outcome Accountability on Judgment Quality" *Organizational Behavior and Human Decision Processes*, 65(1): 1-17.

Simonson, Itamar, and Aimee Drolet. 2004. "Anchoring Effects on Consumers' Willingness-to-Pay and Willingness-to-Accept." *Journal of Consumer Research*, 31(3): 681-690.

Simonson, Itamar, and Barry M. Staw. 1992. “Deescalation Strategies: A Comparison of Techniques for Reducing Commitment to Losing Courses of Action.” *Journal of Applied Psychology*, 77(4): 419-426.

Siniscalchi, Marciano. 2009. “Vector Expected Utility and Attitudes Toward Variation.” *Econometrica*, 77(3): 801-855.

Slonim, Robert, and Alvin E. Roth. 1998. “Learning in High Stakes Ultimatum Games: An Experiment in the Slovak Republic.” *Econometrica*, 66(3): 569-596.

Smith, Vernon L., and James M. Walker. 1993. ‘‘Rewards, Experience and Decision Costs in First Price Auctions.’’ *Economic Inquiry*, 31(2):237-244.

Sobel, Joel. 2005. “Interdependent Preferences and Reciprocity.” *Journal of Economic Literature*, 43(2): 392-436.

Starmer, Chris. 1999. “Experiments in Economics: Should We Trust the Dismal Scientists in White Coats?” *Journal of Economic Methodology*, 6(1): 1-30.

Starmer, Chris, and Robert Sugden. 1993. “Testing for Juxtaposition and Event-Splitting Effects.” *Journal of Risk and Uncertainty*, 6(3): 235-254.

Stott, Henry P. 2006. “Cumulative Prospect Theory's Functional Menagerie.” *Journal of Risk and Uncertainty*, 32(2): 101-130.

Stuhlmacher, Alice F., and Amy E. Walters. 1999. “Gender Differences in Negotiation Outcome: A Meta-Analysis.” *Personnel Psychology*, 52(3): 653-677.

Sugden, Robert. 2003. “Reference-Dependent Subjective Expected Utility.” *Journal of Economic Theory*, 111(2): 172-191.

Sutton, John. 1986. “Non-Cooperative Bargaining Theory: An Introduction.” *Review of Economic Studies*, 53(5): 709-724.

Takemura, Kazuhisa. 1993. "The Effect of Decision Frame and Decision Justification on Risky Choice." *Japanese Psychological Research,* 35(1): 36-40.

Takemura, Kazuhisa. 1994. "Influence of Elaboration on the Framing of Decision." *Journal of Psychology,* 128(1): 33-39.

Telser, Lester G. 1995. "The Ultimatum Game and the Law of Demand." *Economic Journal*, 105(433): 1519-1523.

Tenorio, Rafael, and Timothy N. Cason. 2002. "To Spin or Not To Spin? Natural and Laboratory Experiments from 'The Price is Right'." *Economic Journal*, 112(476): 170-195.

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

Thompson, Leigh. 1990. "Negotiation Behavior and Outcomes: Empirical Evidence and Theoretical Issues." *Psychological Bulletin*, 108(3): 515-532.

Thurstone, Louis L. 1931. "The Indifference Function." *Journal of Social Psychology*, 2(2): 139-167.

Trautmann, Stefan T., Ferdinand M. Vieider, and Peter P. Wakker. 2008. "Causes of Ambiguity Aversion: Known versus Unknown Preferences." *Journal of Risk and Uncertainty,* 36(3): 225-243.

Tversky, Amos, and Daniel Kahneman. 1974. "Judgment under Uncertainty: Heuristics and Biases." *Science*, 185(4157): 1124-1130.

Tversky, Amos, and Daniel Kahneman. 1992. "Advances in Prospect Theory: Cumulative Representation of Uncertainty." *Journal of Risk and Uncertainty*, 5(4): 297-323.

van den Assem, Martijn J., Dennie van Dolder, and Richard H. Thaler. 2012. "Split or Steal? Cooperative Behavior when the Stakes Are Large." *Management Science*, 58(1): 2-20.

van Dolder, Dennie, Martijn J. van den Assem, Colin F. Camerer, and Richard H. Thaler. 2013. "Standing United or Falling Divided? High Stakes Bargaining in a TV Game Show." Working paper. Available at SSRN: http://ssrn.com/abstract=2344422

van Lange, Paul A. M., and D. Michael Kuhlman. 1994. "Social Value Orientations and Impressions of Partner's Honesty and Intelligence: A Test of the Might Versus Morality Effect." *Journal of Personality and Social Psychology*, 67(1): 126-141.

van Lange, Paul A. M., Wilma Otten, Ellen M. N. de Bruin, and Jeffrey A. Joireman. 1997. "Development of Prosocial, Individualistic, and Competitive Orientations: Theory and Preliminary Evidence." *Journal of Personality and Social Psychology*, 73(4): 733-746.

Vieider, Ferdinand M. 2009. "The Effect of Accountability on Loss Aversion." *Acta Psychologica*, 132(1): 96-101.

Vieider, Ferdinand M. 2011. "Separating Real Incentives and Accountability." *Experimental Economics*, 14(4): 507-518.

Vieider,Ferdinand M., Lubomír Cingl, Peter Martinsson and Hrvoje Stojic. 2013. "Separating Attitudes Towards Money From Attitudes Towards Probabilities: Stake Effects and Ambiguity as a Test for Prospect Theory." Working Paper.

Wakker, Peter P. 2010. *Prospect Theory: For Risk and Ambiguity.* Cambridge University Press, Cambridge UK.

Wakker, Peter P., and Amos Tversky. 1993. "An Axiomatization of Cumulative Prospect Theory." *Journal of Risk and Uncertainty*, 7(2): 147-176.

Wakker, Peter P., and Daniel Deneffe. 1996. "Eliciting von Neumann-Morgenstern Utilities When Probabilities Are Distorted or Unknown." *Management Science*, 42(8): 1131-1150.

Walster, Elaine, Ellen Berscheid, and G. William Walster. 1973. "New Directions in Equity Research." *Journal of Personality and Social Psychology,* 25(2): 151-176.

Walters, Amy E., Alice F. Stuhlmacher, and Lia L. Meyer. 1998. "Gender and Negotiator Competitiveness: A Meta-Analysis." *Organizational Behavior and Human Decision Processes*, 76(1): 1-29.

Wedekind, Claus, and Manfred Milinski. 2000. "Cooperation Through Image Scoring in Humans." *Science*, 288(5467): 850–852.

Weigold, Michael F., and Barry R. Schlenker. 1991. "Accountability and Risk Taking." *Personality and Social Psychology Bulletin,* 17(1): 25-29.

West, Stuart A., Ashleigh S. Griffin, and Andy Gardner. 2007. "Evolutionary Explanations for Cooperation." *Current Biology*, 17(16): R661-R672.

Wooldridge, Jeffrey M. 2003. "Cluster-Sample Methods in Applied Econometrics." *American Economic Review*, 93(2): 133-138.

Wu, George, and Alex B. Markle. 2008. "An Empirical Test of Gain-Loss Separability in Prospect Theory." *Management Science*, 54(7): 1322-1335.

Zajonc, Robert B. 1965. "Social Facilitation." *Science,* 149(3681): 269-274.

Summary

This thesis employs natural and laboratory experiments to investigate decision making under risk and uncertainty, cooperative behavior, and bargaining.

Chapter 2 employs data form the British TV game show Golden Balls to study cooperative behavior. In this show, contestants play a variant on the prisoner's dilemma for large and widely ranging stakes averaging over $20,000. In line with previous experimental results, we find evidence that contestants in TV game shows have reciprocal preferences, frame money amounts in relative terms, and tend to stick to their promises. We also find that young males are less cooperative than females and that this gender effect reverses for older contestants as men become increasingly cooperative if age increases.

Chapter 3 uses data from the British TV game show Divided to study bargaining behavior. In this show, contestants bargain over a jackpot that is split into three unequal shares and ranges from about $10,000 to $185,000. In contrast to the commonly held view that fairness concerns will be unimportant when monetary incentives are sufficiently high, we find that individual behavior and outcomes are strongly influenced by equity concerns: those who contributed more to the jackpot claim larger shares, are less likely to make concessions, and take home larger amounts. Threatening to play hardball is ineffective. Although contestants who announce that they will not back down do well relative to others, they do not secure larger absolute amounts and harm others. In addition, there is no evidence of a first-mover advantage and little evidence that demographic characteristics matter.

Together, these two chapters show that many of the patterns found in the experimental laboratory carry over to the drastically different environment of a TV game show. The fact that these findings emerge both in low stakes, relatively anonymous laboratory settings with student subjects as well as in high stakes, public TV game show settings with much

more diverse subject pools, is a positive sign with respect to the generalizability of behavioral findings.

Chapter 4 and 5 investigate the effect of public scrutiny on behavior in both individual choice and interaction choice tasks. Chapter 4 examines whether risky behavior in the limelight differs from that under anonymity. In two experiments, we find that subjects are more risk averse when they make their decisions in the limelight. At the same time, however, their choices follow the same pattern of path dependence in and out of the limelight; subjects take more risk if the game develops either substantially worse or substantially better than expected. As a result, a simple prospect theory model with a path-dependent reference point provides a better explanation for subjects' behavior than a flexible specification of expected utility theory. Additionally, our findings suggest that ambiguity aversion depends on being in the limelight, that passive experience has little effect on risk taking, and that reference points are determined by imperfectly updated expectations.

Chapter 5 examines the effect of social cues, in particular pictures of eyes and pictures of peers, on decisions in both interaction and individual choice tasks. We find that the effect of pictures of eyes is limited to interaction tasks and that it is distinct from the effect of pictures of peers. Whereas pictures of eyes uniformly enhance pro-social behavior in our experiment, this is not the case for pictures of peers. Furthermore, pictures of peers trigger more rational behavior in individual choice tasks that have no moral component, whereas pictures of eyes do not affect behavior in such tasks.

These two chapters suggest that public scrutiny and social cues can influence behavior, and that this influence is not limited to tasks that have a moral component. Together with the first two chapters, this thesis provides a mixed message about the generalizability of findings between different environments. On the one hand, qualitative findings seem to be highly robust across different conditions. In our TV game shows, we observe behavioral patterns that resemble well-documental patterns from the laboratory. Similarly, in Chapter 4 we observe path dependence in risk behavior both in and out of the limelight. On the other hand, however, quantitative estimates appear to be more volatile. These findings are in line with recent studies suggesting that qualitative results generalize between laboratory and field settings even if quantitative results differ (Kagel and Roth, 2000; Tenorio and Cason, 2002; Healy and Noussair, 2004; Isaac and Schnier, 2005; Antonovics, Arcidiacono and Walsh, 2009; Östling et al., 2011; Bolton, Greiner and Ockenfels, 2013).

In Chapter 6, we introduce a new method to measure utility and loss aversion under both risk and uncertainty without the need to introduce simplifying parametric assumptions. Our method extends Wakker en Deneffe's (1996) trade-off method by allowing for standard sequences that include gains, losses, and the reference point. We employ this method to measure utility under both risk and uncertainty, and investigate whether utility takes the same shape for both conditions. This test is critical for models that capture ambiguity aversion through a difference in event weighting between risk and uncertainty, like the multiple priors models and prospect theory.

We cannot reject the hypotheses that both utility and loss aversion are the same for risk and uncertainty, suggesting that utility primarily reflects attitudes towards outcomes. Utility is S-shaped, concave for gains and convex for losses, and there is substantial loss aversion. Our findings support models that explain ambiguity aversion through a difference in event weighting and suggest that descriptive ambiguity aversion models should allow for reference dependence of utility.

Nederlandstalige samenvatting

In dit proefschrift worden laboratoriumexperimenten en natuurlijke experimenten aangewend om risicogedrag, coöperatief gedrag, en onderhandelingsgedrag te bestuderen.

Hoofdstuk 2 gebruikt data van de Britse TV spelshow “Golden Balls” om coöperatief gedrag te onderzoeken. In deze show spelen deelnemers een variant van het welkbekende gevangenendilemma voor grote en sterk variërende geldbedragen, gemiddeld meer dan 13 duizend Britse pond. In lijn met bevindingen uit experimenteel onderzoek vinden observeren we dat mensen wederkerig gedrag vertonen, zich aan hun beloftes houden, en dat deelnemers uitkomsten evalueren ten opzichte van een, normatief gezien, irrelevant referentiepunt. Tevens vinden we dat jonge mannen minder coöpereren dan jonge vrouwen, maar dat dit sekseverschil andersom ligt voor oudere deelnemers doordat mannen zich coöperatiever gedragen naarmate ze ouder zijn.

Hoofdstuk 3 gebruikt data van de Britse TV spelshow “Divided” om onderhandelingsgedrag te analyseren. In deze show onderhandelen deelnemers over een jackpot die in drie ongelijke delen wordt gesplitst. Het betreft ook hier hoge bedragen: de jackpot is gemiddeld meer dan 33 duizend Britse pond. In tegenstelling tot het veelgehoorde punt dat rechtvaardigheidsoverwegingen niet van belang zullen zijn als monetaire prikkels hoog genoeg zijn, vinden we dat individueel gedrag en uitkomsten sterk worden beïnvloed door de mate waarin spelers hebben bijgedragen aan de jackpot: bij een hogere bijdrage claimt men een groter deel van de jackpot, is men minder geneigd concessies te doen, en neemt men een groter deel van de jackpot mee naar huis. Daarnaast vinden we dat het aankondigen van een agressieve onderhandelingsstrategie geen beter resultaat oplevert. Hoewel deelnemers die stellig aankondigen geen enkele concessie te zullen doen met een relatief groot deel van de jackpot naar huis gaan, incasseren zij geen groter bedrag in absolute zin en heeft hun gedrag een negatieve impact voor anderen. Tot slot is er geen indicatie dat het voordelig is als eerste een claim op tafel te kunnen leggen, en zijn er weinig aanwijzingen dat demografische factoren verklarende kracht hebben.

Tezamen geven deze twee hoofdstukken aan dat veel patronen uit het experimentele laboratorium generaliseren naar de drastisch verschillende omgeving van een TV spelshow. Het feit dat deze bevindingen uit de uiterst publieke omgeving van een TV spelshow, met hoge geldbedragen en een diverse groep "proefpersonen" stroken met bevindingen uit experimenten waar bedragen doorgaans laag zijn, studenten als proefpersoon fungeren en beslissingen een hoge mate van anonimiteit hebben, is een positief teken wat betreft de generaliseerbaarheid van bevindingen omtrent beslissingsgedrag.

Hoofdstuk 4 en 5 onderzoeken het effect van bekeken worden op beslissingsgedrag, zowel waar het individueel gedrag als interactiegedrag betreft. Hoofdstuk 4 onderzoekt of risicogedrag anders is als mensen hun keuzes maken voor een groot publiek (in de schijnwerpers staan) dan wanneer ze hun keuzes maken onder meer anonieme omstandigheden. In twee experimenten vinden we dat proefpersonen minder risico nemen wanneer ze in de schijnwerpers staan. Tegelijkertijd volgen de beslissingen een vergelijkbaar patroon van padafhankelijkheid: proefpersonen nemen meer risico als het spel zich substantieel beter of slechter ontwikkelt dan oorspronkelijk verwacht. Een eenvoudig *prospect theory* model met een padafhankelijk referentiepunt verklaart het gedrag van proefpersonen beter dan een flexibele specificatie van verwachte-nutstheorie. Tevens vinden we dat ambiguïteitsaversie sterker is als anderen meekijken, dat passieve ervaring met het spel weinig effect heeft op risicogedrag, en dat referentiepunten worden bepaald door onvolledige aanpassing van verwachtingen.

Hoofdstuk 5 onderzoekt het effect van subtiele sociale stimuli op beslissingen in interactieve en individuele keuzesituaties, in het bijzonder van foto's van ogen en foto's van medestudenten. We vinden dat het effect van foto's van ogen zich beperkt tot interactietaken en dat dit effect verschilt van dat van foto's van medestudenten. Waar foto's van ogen sociaal gedrag stimuleren in interactietaken, is dit niet altijd het geval voor foto's van medestudenten. Daarnaast zorgen foto's van medestudenten voor meer rationeel gedrag in individuele taken zonder morele component. Foto's van ogen hebben geen invloed in dergelijke taken.

Deze twee hoofdstukken laten zien dat bekeken worden en zelfs subtiele sociale stimuli gedrag kunnen beïnvloeden, en dat deze invloed zich niet beperkt tot taken met een morele component. Samen met de eerste twee hoofdstukken schetst dit proefschrift een gemengd beeld wat betreft de generaliseerbaarheid van bevindingen tussen verschillende omgevingen. Aan de ene kant lijken kwalitatieve patronen robuust over verschillende condities. In onze spelshows vinden we gelijksoortige gedragingen als voorheen geobserveerd in het gedragslaboratorium. Tevens zien we in hoofdstuk 4 dat de padafhankelijkheid van risicogedrag niet anders is wanneer men in de schijnwerpers staat. Tegelijkertijd echter, blijken kwantitatieve resultaten minder robuust. Deze bevindingen bevestigen de uitkomsten van recente studies die laboratorium- en veldgedrag met elkaar vergelijken (Kagel and Roth, 2000; Tenorio and Cason, 2002; Healy and Noussair, 2004; Isaac and Schnier, 2005; Antonovics, Arcidiacono and Walsh, 2009; Östling et al., 2011; Bolton, Greiner and Ockenfels, 2013).

In hoofdstuk 6 introduceren we een nieuwe methode om de nutsfunctie en verliesaversie te meten onder zowel risico en onzekerheid. Onze methode vereist geen simplificerende parametrische assumpties en bouwt voort op de *trade-off* methode van Wakker en Deneffe (1996) door het mogelijk te maken standaardreeksen op te zetten die zowel winsten, verliezen en het referentiepunt bevatten. We gebruiken deze methode om nut te meten onder risico en onzekerheid en onderzoeken vervolgens of de nutsfunctie verschilt tussen deze condities. Deze test is cruciaal voor modellen die ambiguïteitsaversie verklaren door verschillen in het wegen van gebeurtenissen tussen risico en onzekerheid, zoals het *multiple priors* model en *prospect theory*.

We zijn niet in staat de hypotheses te verwerpen dat de nutsfunctie en verliesaversie hetzelfde zijn onder risico en onzekerheid. Dit resultaat suggereert dat de nutsfunctie puur attitudes ten opzichte van uitkomsten betreft. We vinden dat de nutsfunctie S-vormig is, concaaf voor winsten en convex voor verliezen, en dat er sprake is van aanzienlijke verliesaversie. Onze bevindingen ondersteunen modellen die ambiguïteitsaversie verklaren aan de hand van een verschil in de weging van gebeurtenissen. Daarnaast suggereren onze bevindingen dat descriptieve modellen voor ambiguïteitsaversie rekening moeten houden met het feit dat mensen uitkomsten evalueren met betrekking tot een referentiepunt.

About the author

Dennie van Dolder (1984) holds a Bachelor's degree in Sociology and a Research Master's degree in Sociology and Social Research from Utrecht University, and a Master's degree in Behavioural Economics from the University of Nottingham (United Kingdom), all with highest honors. He worked on his PhD thesis at Erasmus University Rotterdam under the supervision of Peter Wakker and Han Bleichrodt. As part of his PhD training, Dennie spent time as a visiting researcher at the University of Chicago Booth School of Business. Currently, he works as a research fellow at the University of Nottingham, where he is affiliated with the Centre for Decision Research and Experimental Economics (CeDEx) and the ESCR funded Network for Integrated Behavioural Sciences (NIBS). He has published articles in Management Science, Evolution and Human Behavior, and PLoS ONE.

The Tinbergen Institute is the Institute for Economic Research, which was founded in 1987 by the Faculties of Economics and Econometrics of the Erasmus University Rotterdam, University of Amsterdam and VU University Amsterdam. The Institute is named after the late Professor Jan Tinbergen, Dutch Nobel Prize laureate in economics in 1969. The Tinbergen Institute is located in Amsterdam and Rotterdam. The following books recently appeared in the Tinbergen Institute Research Series:

539 J.A. ROSERO MONCAYO, *On the importance of families and public policies for child development outcomes*
540 E. ERDOGAN CIFTCI, *Health Perceptions and Labor Force Participation of Older Workers*
541 T.WANG*, Essays on Empirical Market Microstructure*
542 T. BAO, *Experiments on Heterogeneous Expectations and Switching Behavior*
543 S.D. LANSDORP*, On Risks and Opportunities in Financial Markets*
544 N. MOES, *Cooperative decision making in river water allocation problems*
545 P. STAKENAS, *Fractional integration and cointegration in financial time series*
546 M. SCHARTH*, Essays on Monte Carlo Methods for State Space Models*
547 J. ZENHORST, *Macroeconomic Perspectives on the Equity Premium Puzzle*
548 B. PELLOUX, *the Role of Emotions and Social Ties in Public On Good Games: Behavioral and Neuroeconomic Studies*
549 N. YANG, *Markov-Perfect Industry Dynamics: Theory, Computation, and Applications*
550 R.R. VAN VELDHUIZEN*, Essays in Experimental Economics*
551 X. ZHANG, *Modeling Time Variation in Systemic Risk*
552 H.R.A. KOSTER, *The internal structure of cities: the economics of agglomeration, amenities and accessibility.*
553 S.P.T. GROOT, *Agglomeration, globalization and regional labor markets: micro evidence for the Netherlands.*
554 J.L. MÖHLMANN, *Globalization and Productivity Micro-Evidence on Heterogeneous Firms, Workers and Products*
555 S.M. HOOGENDOORN, *Diversity and Team Performance: A Series of Field Experiments*
556 C.L. BEHRENS, *Product differentiation in aviation passenger markets: The impact of demand heterogeneity on competition*
557 G. SMRKOLJ, *Dynamic Models of Research and Development*
558 S. PEER, *The economics of trip scheduling, travel time variability and traffic information*
559 V. SPINU, *Nonadditive Beliefs: From Measurement to Extensions*

560 S.P. KASTORYANO, *Essays in Applied Dynamic Microeconometrics*
561 M. VAN DUIJN, *Location, choice, cultural heritage and house prices*
562 T. SALIMANS, *Essays in Likelihood-Based Computational Econometrics*
563 P. SUN, *Tail Risk of Equidity Returns*
564 C.G.J. KARSTEN, *The Law and Finance of M&A Contracts*
565 C. OZGEN, *Impacts of Immigration and Cultural Diversity on Innovation and Economic Growth*
566 R.S. SCHOLTE, *The interplay between early-life conditions, major events and health later in life*
567 B.N.KRAMER, *Why don't they take a card? Essays on the demand for micro health insurance*
568 M. KILIÇ, *Fundamental Insights in Power Futures Prices*
569 A.G.B. DE VRIES, *Venture Capital: Relations with the Economy and Intellectual Property*
570 E.M.F. VAN DEN BROEK, *Keeping up Appearances*
571 K.T. MOORE, *A Tale of Risk: Essays on Financial Extremes*
572 F.T. ZOUTMAN, *A Symphony of Redistributive Instruments*
573 M.J. GERRITSE, *Policy Competition and the Spatial Economy*
574 A. OPSCHOOR, *Understanding Financial Market Volatility*
575 R.R. VAN LOON, *Tourism and the Economic Valuation of Cultural Heritage*
576 I.L. LYUBIMOV, *Essays on Political Economy and Economic Development*
577 A.A.F. GERRITSEN, *Essays in Optimal Government Policy*
578 M.L. SCHOLTUS, *The Impact of High-Frequency Trading on Financial Markets*
579 E. RAVIV, *Forecasting Financial and Macroeconomic Variables: Shrinkage, Dimension reduction, and Aggregation*
580 J. TICHEM, *Altruism, Conformism, and Incentives in the Workplace*
581 E.S. HENDRIKS, *Essays in Law and Economics*
582 X. SHEN, *Essays on Empirical Asset Pricing*
583 L.T. GATAREK, *Econometric Contributions to Financial Trading, Hedging and Risk Measurement*
584 X. LI, *Temporary Price Deviation, Limited Attention and Information Acquisition in the Stock Market*
585 Y. DAI, *Efficiency in Corporate Takeovers*
586 S.L. VAN DER STER, *Approximate feasibility in real-time scheduling: Speeding up in order to meet deadlines*
587 A. SELIM, *An Examination of Uncertainty from a Psychological and Economic Viewpoint*
588 B.Z. YUESHEN, *Frictions in Modern Financial Markets and the Implications for Market Quality*