AMBIGUITY IN SOCIAL DILEMMAS

ISBN 978 90 3610 474 6

Cover design: Crasborn Graphic Designers bno, Valkenburg a.d. Geul

This book is no. **684** of the Tinbergen Institute Research Series, established through cooperation between Rozenberg Publishers and the Tinbergen Institute. A list of books which already appeared in the series can be found in the back.

Ambiguity in Social Dilemmas

Ambiguïteit in sociale dilemma's

Thesis

to obtain the degree of Doctor from the
Erasmus University Rotterdam
by command of the
rector magnificus
Prof.dr. H.A.P. Pols

and in accordance with the decision of Doctoral Board.
The public defense shall be held on
Thursday January 5, 2017 at 13:30 hours

by
UYANGA TURMUNKH
born in Ulaanbaatar, Mongolia

Erasmus University Rotterdam

Doctoral Committee:

Promotors:

Prof.dr. P.P. Wakker

Prof.dr. M.J. van den Assem

Other members:

Prof.dr. A. Baillon

Prof.dr. H. Bleichrodt

Dr. A. Perea

Acknowledgment

I have been fortunate to become a student of my advisor Peter P. Wakker. For four years I benefited by observing and learning from Peter's reasoning, speech, and scholarly integrity. I thank him first and foremost, for I am a better thinker, researcher, and person because of his kind mentorship.

I thank my second advisor and first coauthor Martijn J. van den Assem, for giving me the opportunity to work on a fun and challenging project, which jump-started my research at the start of my PhD when I was still struggling to chart an independent course. Through the project with Martijn, I got a chance to work with my other coauthor Dennie van Dolder, whose confident work on our project, coming from a young and newly minted researcher, served as a special personal inspiration for me. I was also inspired by, and learned much from another young researcher, Chen Li, who is brave, tenacious, and who can move things at incredible speed. I want to extend special thanks to another coauthor, my friend, and my officemate of three years, Tong Wang, who, being my peer and friend, had to endure the uninhibited display of my shortcomings.

My development as a researcher has benefited greatly from my visit to the University of Michigan. I want to thank Richard Gonzalez for hosting me at Michigan and for his kind support. I want to thank Erasmus Trustfonds for the travel grant, which made the research visit possible. I was able to grow also because I was surrounded by excellence at the Behavioral Economics group of Erasmus University—people producing quality output, thinking boldly, arguing bravely, sharing wisdom, betting, brain teasing, laughing, becoming vegan, and running. I thank all my colleagues: Han, Kirsten, Aurélien, Arthur, Rogier, Amit, Vitalie, Jingni, Paul, Aysil, Zhenxing, Ning, Yu, Zhihua, Jan S., Jan H., and Georg. My work (and life) at Erasmus University would have been half as pleasant if not for the spirited, warm, and supportive presence of Judith, Carolien, Arianne, Ine, and Mirjam.

I want to thank three special individuals, whose friendship filled the void that I, just like anyone moving to a new place, carried when I started my PhD in Rotterdam. I thank my friends Violeta Misheva, Zara Sharif, and Ilke Aydogan. Life in the past four years has been grounded, rich, and heartfelt because of them. I also want to thank a group of people who in one way or another added warmth and quality to my life in the Netherlands: Hao, Lydia, Jonathan, Aaron, Alex, Mehmet, Sander, Marcin, Sasha, Uwe, Max, Olivier, Sait, Eszter, Gergely, Tomasz, Mark, Arturas, Lisette, Dasha, Anghel, Piotr, Simin, and Jindi (aka Jimmy).

I want to thank Professors Andreas Irmen, Kerry Whiteside, Brian A'Hearn, Judith Mueller, and Arnold Feldman for making life more interesting. They are partly the reason for why I consider teaching a noble profession. I want thank my friend Frank, whom I idolize unabashedly and whose presence in my life has resulted in countless improvements in my health, wealth, and well-being. I thank Evgeny, who has been a source of happiness and who taught me the value of time. I thank my friends Jana, Marty, Pavel, Marina, Thomas, and Simon, who from afar and without their knowledge have filled me with warmth all this while. I thank Kevin, though I will never be able to thank him enough. I am happy that he, Dawn, Renee, and Charlie are in my life. I am happy because I have a sister, who is smarter and a better human than I am. I thank my father, my spirit animal, for being my father. I thank my mother for being.

Contents

Chapter 1

Introduction

Much of our collective thriving as a society relies "merely" on each individual one of us pursuing our own best interests. Even when this pursuit of self-interest pits some of us against others so that the resulting competition inevitably leads to losses for the losers, the society as a whole often gains, by discovering, employing, and consuming the winners' ideas. There are, however, situations, in which this fortuitous alignment between individual self-interest and collective well-being does not exist. In such situations called *social dilemmas* an individually rational choice of each member turns out paradoxically to be collectively irrational.

Some of our most consequential collective decisions are hampered by social dilemmas. Overthrowing an oppressive regime, for instance, is a collectively rational decision that requires individual sacrifices from many participants in the movement. While success is immensely desirable, each individual citizen contemplating the decision to join or not must trade off the increase in the chance of success due to own participation against the possible loss, in the best case, of the foregone value of time spent elsewhere and, in the worst case, of one's personal freedom or even life. If not many others choose to join, the chance of success will be low and the losses to the participating few grave, and so from an individual point of few it would be rational not to join. If on the other hand many do join, success will be likely and the losses benign, but from an individual point of view the rational (and tempting) choice would still be not to join, so that one can avoid—however benign—the personal losses while still benefiting from the success of the movement. And yet, if every citizen makes the rational choice of not joining, all suffer under continued oppression.

Perhaps less dramatic than social movements, but no less consequential, is the challenge of managing our common resources. Here again, collective interest dictates moderation, while individual rationality tempts us toward excess. Mitigating the threat of climate change, possibly the most consequential decision facing the world, requires sovereign nations to sacrifice individual growth in the effort to avert a collective disaster. Whether large or small, social dilemmas share the property that in the absence of some external enforcement of cooperation, the individual decision makers are always tempted (though by no means doomed) to defect.

To be sure, there are those who would, regardless of others' actions, choose the individually advantageous course in any social dilemma. The premise of this thesis, however, is that, within reasonable ranges of distance between potential gains and losses, most people are conditionally cooperative; that is, they are willing to choose the collectively advantageous course if only sufficiently many others would do the same. For conditionally cooperative individuals the decision to cooperate or defect crucially depends on their judgment of how others will behave. Yet, others' behavior is often uncertain to us. We must instead rely on our past experiences in similar interactions or with similar others and the verbal and non-verbal cues of the interacting others to make our best guess about how they will behave. The uncertainty that people face in a social dilemma about others' behavior, how it affects their choices, and how it can be reduced is the subject of this thesis.

In economics, (behavioral) game theory studies social dilemmas by means of stylized models called *games* designed to capture the basic incentive structures, which decision makers face in commonly occurring social dilemmas. Traditional analyses of games invariably assumed that all uncertainties can be modeled as *risk*, that is, they assumed that objective probabilities of others' uncertain choices are known. In reality these probabilities are rarely known, and decisions in games are made under *ambiguity*—uncertainty with no objective probabilities available. In the literature on individual decision-making, it is well known that people treat ambiguity in a fundamentally different way than risk. Yet, in many experimental studies of games people's beliefs about others' choices are measured under the traditional assumption, which ignores ambiguity attitudes.

Chapter 2 of this thesis extends a method for measuring ambiguity attitudes for uncertain events in individual choice to game situations. Unlike the measurements of beliefs in experimental games made under the behaviorally invalid assumption that subjects are ambiguity neutral, our method can correct for ambiguity attitudes and thus give accurate measures of beliefs even if the assumption of ambiguity neutrality is violated. We use our technique to investigate the role of ambiguity in trust games. Previous studies sought to understand the role of uncertainty about others' trustworthiness by measuring risk attitudes, and finding no relation. We show that ambiguity attitudes do matter: ambiguity aversion reduces trusting decisions. Moreover, ambiguity corrected beliefs are found to contribute to the decision to trust. We are also able to confirm, on the basis of revealed preference data, that introspective survey questions about trust such as those used by the World Values Survey, which were found to not correlate well with experimental measures of trusting decisions, do capture trust in the sense of belief in trustworthiness of others.

Chapter 3 investigates betrayal aversion. When it comes to dealing with uncertainty about others' trustworthiness, it has been reported that people are betrayal averse. They experience extra aversion towards subjecting themselves to the possibility of betrayal by a fellow human than they would towards the possibility of a bad outcome resulting from unfavorable forces of nature. Studies reporting betrayal aversion sought to study the human vs. nature divide in people's uncertainty preferences within the risk domain. That is, they sought to compare willingness to take risk in a situation where the probabilities of monetary outcomes were generated by choices made by opponents to a situation where the same probabilities of the same outcomes were generated by a chance device. Any difference was then attributed to the additional averseness that people experienced when outcomes resulted from interactions with others. Chapter 3 of this thesis shows that the findings of these studies can be explained by ambiguity aversion rather than betrayal aversion. When ambiguity attitudes are controlled for, we find no systematic betrayal aversion. People in fact display less preference for nature-generated ambiguity than betrayal ambiguity.

Chapter 4 considers social dilemma situations where conditionally cooperative people would like to accurately predict others' intended actions. Often, the information that is available for predicting what others will do is limited to non-binding and non-verifiable communication. Under the traditional assumption that lying is costless, such cheap talk is not informative when incentives are insufficiently aligned, as is usually the case in social dilemmas. Chapter 4 of this thesis uses data from a TV game show to investigate the credibility of pre-play cheap talk in a game that resembles the classical prisoner's dilemma. In each episode of the game show studied in this chapter, two contestants simultaneously decide to either split (cooperate) or steal (defect) a sum of money that on average exceeds $20,000. Prior to their decisions, the contestants briefly engage in a free-form discussion about the choice at hand. During the talk, they typically exchange multiple statements, most of which involve giving or eliciting some type of signal that the intended decision is to split.

We investigate whether the distinction between different types of statements adds to the predictive power of the contestants' cheap talk. In this chapter, we build on insights from psychology to argue that lies are less costly if they are malleable to ex-post reinterpretation as truths. We propose a typology of statements that distinguishes them according to two dimensions. First, our typology discriminates between statements explicitly expressing that the contestant will choose split and statements that only implicitly signal that the contestant will do so. Second, it discriminates between unconditional statements and statements that carry an element of conditionality on what the other will do. Our hypothesis is that people who defect prefer to make statements that allow them to deny the fact that they are lying. Our data show that malleability is indeed a plausible criterion for judging the credibility of cheap talk.

Chapter 5 reports results of a field experiment conducted at Erasmus University, in collaboration with the university's student affairs office. Behavioral economics has recently enjoyed much attention from policy-makers. In 2010, for instance, the government of the United Kingdom set up the world's first government institution dedicated to applying insights from the behavioral sciences to public policy and services. More recently, in the United States President Obama has established the U.S. government's own "nudge" unit. As the dubbing of such agencies suggests, much of the popularity of behavioral economics among policy-makers has to do with the findings from research to date that incorporating behavioral nudges into policy design can improve its efficacy and reach, and that this can be done inexpensively. Messaging nudges have been particularly popular as such interventions can be implemented using existing government databases easily and virtually at no additional cost.

The goal of our experiment was to test some of the most popular messaging nudges, on whether these can influence students at Erasmus University to increase their participation rate in the online evaluations of their courses. Participation in the evaluation surveys is voluntary at Erasmus University. Moreover, filling out a course evaluation can be seen as the decision to contribute to a public good: costs on students' time are private while potential benefits in the form of improved future course quality are public, and in this case, unlikely for the participating student to share in the benefit. Unsurprisingly, participation rates have been extremely low.

Our experiment randomly assigned all registered students to one of three messaging nudges. The treatment messages emphasized the real impact of course evaluations, descriptive norms, and commitment, respectively. The results are puzzling in that we find opposite effects of our nudges on seemingly similar groups. All three nudges helped increase the participation rate among Master students, while depressing the participation of Bachelor students. Some observable differences partly close the gap, but not entirely. These results echo the numerous other experiments conducted by the aforementioned “nudge” units using similar messaging interventions, which find no effects but are not published in scientific journals. Our results emphasize the need to publish the mixed findings of nudging interventions so as to further our understanding of the context and target dependence of their effects.

This thesis was built on the premise that, faced with a social dilemma situation, within reasonable limits most people are willing to cooperate, provided that most others also cooperate. The questions that it raised were motivated by the desire to understand better the circumstances under which such conditionally cooperative people, while willing, may fail to cooperate. Using laboratory experiments, we investigated the role that uncertainty about others’ intentions plays in such decisions. Using data from a TV game show, we studied whether words can be relied on as signals for judging others’ intentions more accurately. Finally, through a field experiment, we tested the possibility of nudging people towards more cooperation.

Chapter 2

Ambiguity and Trust in Games

with CHEN LI AND PETER P. WAKKER

2.1 Introduction

Keynes (1921) and Knight (1921) emphasized the importance of developing models for ambiguity (unknown probabilities), given its ubiquity in economic decisions and everyday life. Ellsberg (1961) showed that such models have to be fundamentally different from traditional risk models. Despite the importance of ambiguity, it was not until the end of the 1980s that people succeeded in developing the first decision models (Gilboa 1987; Gilboa and Schmeidler 1989; Schmeidler 1989). Since then, many fields in economics started to catch up with ambiguity. One such field is game theory.

In games, people make decisions under uncertainty, where a major source of uncertainty originates from opponents' strategy choices. Although traditional analyses of games invariably assumed that all uncertainties can be expressed in terms of probabilities, in reality these probabilities are usually unknown. With the increase in awareness of ambiguity, many theoretical studies have applied ambiguity models to analyses of games, producing more realistic predictions about people's choices. However, experimental exploration of the ambiguous nature of games is lagging behind. Section 2.5 gives references. For instance, many experimental studies measure subjective beliefs of players[1] but they commonly assume such beliefs to be Bayesian (ambiguity neutral) additive probabilities, ignoring ambiguity attitudes. Exceptions are Ivanov (2011) and Bellemare, Kroger, and van Soest (2008). Even if one assumes that using such probabilities is rational[2], then still this assumption does not hold empirically.

[1] See for instance Armantier and Treich (2009), Blanco et al. (2010), Costa-Gomes and Weizsäcker (2008), Heinemann, Nagel, and Ockenfels (2009a, b), Huck and Weizsäcker (2002), Neri (2015), Nyarko and Schotter (2002), Palfrey and Wang (2009), Rutström and Wilcox (2009), and Trautmann and van de Kuilen 2015 footnote 16).
[2] This deviates from the rationality judgments by Ellsberg (1961), Gilboa et al. (2010), and others.

The first difficulty in applying ambiguity theories to natural events as occurring in experimental games arises from the necessity to control for beliefs in ambiguity measurements. Therefore, the literature on ambiguity measurements has so far focused on artificial ambiguity, where control of beliefs is possible using symmetries introduced by experimental designs (using Ellsberg urns or experimenter-specified probability intervals). Such symmetries are not available for natural ambiguous events, including moves of opponents in strategic conflicts. Baillon et al. (2016) resolved this first difficulty. They introduced a new ambiguity measurement method that works for all natural events without the need of artificial symmetries in beliefs.

A second difficulty in applying ambiguity theories to experimental games comes from the complication of potential strategic interactions of ambiguity measurements with the games, which was not an issue in individual decision making situations where most ambiguity measurements have been applied so far. We adapt Baillon et al.'s (2016) method to game theory by implementing side bets and their incentivization in a way that matches players to opponents while avoiding strategic interactions. Thus we resolve the second difficulty.

We use our technique to investigate the role of ambiguity in trust games. The role of trust received much interest in behavioral game theory (for overview, see Camerer 2003 chapter 2; Fehr 2009; Johnson and Mislin 2011), but no controls for ambiguity were available yet. Attitudes toward uncertainty matter for trust decisions because it is uncertain whether one's trust will be reciprocated. Previous studies focused on how people's risk[3] attitude contributes to their trust decisions, but found no relation (Eckel and Wilson 2004; Houser et al. 2010). However, the decision to trust usually is not a decision under risk, but rather under ambiguity. That is, we rarely know an objective probability of others being trustworthy. It has been well documented in the literature that people treat ambiguity differently than risk (Ellsberg 1961; Trautmann and van de Kuilen 2015). Hence, to better understand people's trust decisions, it is desirable to analyze these as decisions under ambiguity. To illustrate, assume that we observe that person A decides to trust whereas B does not. Then it is unclear whether A is more trusting (believing more that the other is trustworthy) or, instead, less ambiguity averse. Using Baillon et al.'s (2016) measurement method, we can separate people's subjective beliefs from their ambiguity attitudes towards others' trustworthiness.

[3] In this paper, risk refers to known probabilities, ambiguity to unknown probabilities, and uncertainty encompasses both.

Baillon et al. (2016) introduced two indices capturing two components of people's ambiguity attitudes: an aversion index and an insensitivity index. In the context of the trust decision the ambiguity aversion index captures how much people dislike the ambiguity about their trustee's trustworthiness. The insensitivity index captures how much people perceive and understand ambiguity in the decision situation: the more ambiguity they perceive, the more they treat all events alike, resulting in insensitivity (lower discriminatory power) towards likelihood levels. As regards beliefs, we extrapolated people's beliefs from their choices while correcting for their ambiguity attitudes. This makes our belief measurements properly incentivized without the distortion of ambiguity attitudes.

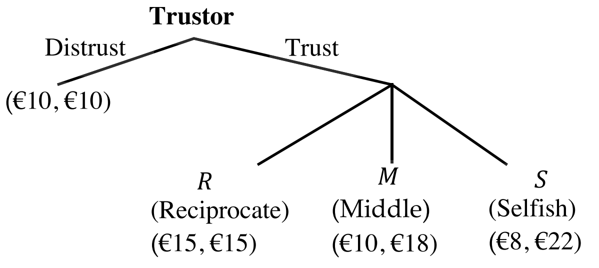
We find more ambiguity averse people to be less likely to trust, whereas insensitivity plays no role in this decision. Further, we confirm that people who have more optimistic beliefs about others' trustworthiness are more likely to trust others. These results are psychologically plausible. We can also shed new light on several open questions in the literature. We confirm, on the basis of revealed preference data, that introspective survey questions such as in the World Values Survey (WVS) do capture trust in the commonly accepted sense of belief in trustworthiness of others (Gambetta et al. 2000). We can also confirm a proper signaling role of own type about population composition, which is a rational form of what Ross, Greene, and House (1977) called false consensus. For rational versions see Dawes (1990) and Prelec (2004).

2.2 Method

Figure 2.1 shows the trust game we used. A trustor faces a binary choice. If she chooses the option distrust, both she and her trustee receive €10 for sure and there is no uncertainty. She can also choose a trust option, whose outcome is uncertain. Then how much she receives is up to the trustee's choice from three allocation options, R (Reciprocate) = (€15, €15), M (Middle) = (€10,€18), and S (selfish) = (€8, €10). Here the first amount is the payment for the trustor and the second is for the trustee.

The game we used is a modification of the trust game used by Bohnet and Zeckhauser (2004). The only difference is that the trustee has one extra option (M) to choose from. Option M gives the trustee the alternative to be selfish without hurting the trustor but at a slight efficiency cost—the total payment is then €28 instead of €30. We added this option because to measure people's ambiguity aversion and insensitivity while controlling for their beliefs, we need at least three events (Baillon et al. 2016).

FIGURE 2.1 Trust game



Let E_i ($i = r, m,$ or s) denote the event that the trustee chooses option I ($I = R, M,$ or S). These events are exhaustive and mutually exclusive. We refer to them as *single events*. A *composite event*, denoted E_{ij} ($j \neq i$), is the union $E_i \cup E_j$ of two single events. For each event E (E_i or E_{ij}) and a fixed outcome $X > 0$ ($X =$ €15 in the experiment), $X_E 0$ denotes a, possibly ambiguous, prospect that pays X if event E happens and 0 otherwise. Similarly, $X_q 0$ denotes a risky prospect that pays X with probability q and 0 with probability $1 - q$.

DEFINITION 2.1 The *matching probability* m (m_i or m_{ij}) of an event E (E_i or E_{ij}) is the probability such that the decision maker is indifferent between prospects $X_E 0$ and $X_m 0$.

The matching probability m of an event E reflects the decision maker's subjective belief in event E, but distorted by her ambiguity attitude. Dimmock, Kouwenberg, and Wakker (2016 Theorem 3.1) showed that, if we know beliefs, then matching probabilities capture people's ambiguity attitudes while controlling for their risk attitudes. Baillon et al. (2016) added the control for beliefs. Next, we briefly introduce the two indices of Baillon et al. (2016) that we use. Let $\overline{m_s} = (m_r + m_m + m_s)/3$ denote the average single-event matching probability and let $\overline{m_c} = (m_{rm} + m_{rs} + m_{ms})/3$ denote the average composite-event matching probability.

DEFINITION 2.2 The a*mbiguity aversion index* is

$$b = 1 - \overline{m_s} - \overline{m_c}. \qquad (2.1)$$

DEFINITION 2.3 The a*(mbiguity-generated)-insensitivity index* is

$$a = 3 \times \left(\frac{1}{3} - (\overline{m_c} - \overline{m_s})\right). \qquad (2.2)$$

Under ambiguity neutrality, $\overline{m_s} = \frac{1}{3}$ and $\overline{m_c} = \frac{2}{3}$, so that both indices are 0. For an ambiguity averse person the matching probabilities are low and the aversion index accordingly is high. She is willing to pay a premium (in winning probability) to avoid ambiguity. A maximally ambiguity averse person has all matching probabilities 0 and the aversion index is 1. For ambiguity seeking subjects, the aversion index is negative.

The insensitivity index concerns the (lack of) discriminatory power of the decision maker regarding different levels of likelihood. For a completely insensitive person who does not distinguish between composite and single events, $\overline{m_c} = \overline{m_s}$, and the insensitivity index takes its maximal value 1. This happens for people who take all uncertainties as fifty-fifty. The better a person discriminates between composite and single events, the larger $\overline{m_c} - \overline{m_s}$ is and the smaller the insensitivity index is. The index captures perception of ambiguity. The more ambiguity a person perceives, the more all events are perceived as one blur and the higher the index is. The index also captures cognitive discriminatory power.

Theoretical studies have focused on ambiguity aversion, but empirical studies have shown that insensitivity is also important (Trautmann and van de Kuilen 2015). Baillon et al. (2016) showed that the indices are a common generalization of many existing ambiguity aversion indices in the literature, proposed under various ambiguity theories.[4]

Our elicitation method also allows for extrapolating a-neutral probabilities p_i. These can be interpreted as the beliefs of an ambiguity neutral twin of the decision maker, who is exactly the same as the decision maker except that she is ambiguity neutral. That is, a-neutral probabilities are additive subjective probabilities that result after correcting for ambiguity attitudes. Baillon et al. (2016) showed that, under certain assumptions:

$$p_i = \frac{3(\overline{m_c} - \overline{m_s}) + 3m_i - 3m_{jk} + 2(1-a)}{6(1-a)}, \text{ where } \{i, j, k\} = \{r, m, s\}. \qquad (2.3)$$

[4] These include Schmeidler (1989) and Dow and Werlang (1992) for Choquet expected utility, Abdellaoui et al. (2011) and Dimmock, Kouwenberg, and Wakker (2016) for prospect theory, and Dimmock et al. (2015) and Epstein and Schneider (2010) for multiple priors.

2.3 Experimental design

Subjects

In total, 182 subjects (38 in the first wave and 144 in the second wave) were recruited from the subject pool of the experimental laboratory at Erasmus School of Economics. 56% were male.

FIGURE 2.2 Trust game: trustor decision situation

The following may be inside your envelope.

Recall that you are matched with one other participant. You can instruct the experimenters to give you one of the following two options:

Option 1: Follow your partner's instruction for payment

Option 2: Pay **€10** to each of you

If you instruct the experimenters to give you Option 1,

your partner's instruction will determine the payments for the two of you. Your partner can instruct the experimenters to give you one of the following three options:

Option A: Pay **€15** to each of you;

Option B: Pay you **€10**, pay him/her **€18**;

Option C: Pay you **€8**, pay him/her **€22**.

So if your partner has instructed to give Option A, you and your partner will get €15 each. If your partner has instructed to give Option B, you will get €10 and your partner €18. Finally, if your partner has instructed to give Option C, you will get €8 and your partner €22.

If you instruct the experimenters to give you Option 2,

you and your partner will get €10 each (and your partner's instruction will play no role).

Incentives

The experiment was computer-based[5] consisted of seven sessions, and was incentivized using the prior incentive system (Prince; Johnson et al. 2015). At the beginning of each session (with n subjects), one volunteer was invited to randomly draw $n/2$ pairs of sealed

[5] Instructions and the online experiment can be found at http://www.peterwakker.com/trustnew/begin.php. For testing, use any 4-digit subject ID's starting with 6 (e.g. 6067) to go through the experiment.

envelopes. Each subject would then draw one envelope from the pile selected by the volunteer.

It was explained to each subject that, throughout the experiment, she would be paired with a partner whose subject ID was inside the envelope. During the experiment, she would face different decision situations, where her payments depended on both her own and her partner's decisions. One of these decision situations was inside the envelope, and this was the only one that mattered for the real payment at the end.

Stimuli

During the experiment, there were three types of decision situations. Subjects also answered some demographic and general survey questions, which were not incentivized. Each subject first faced the trustor decision in the trust game (Figure 2.2). It was explained to her that her own and her partner's choice as a trustee would be used to determine their final payment if this decision situation came out of her envelope.

After making their choices as the trustor, subjects proceeded to the next part of the experiment, where they faced 24 decision situations designed to elicit their matching probabilities. Figure 2.3 depicts a typical decision situation. In the decision situation depicted in the figure a subject chose between two options, both of which might pay her €15 but under different conditions. Option 1 was an ambiguous prospect paying €15 if her partner (as the trustee) chose option R in the trust game. Option 2 was a risky prospect paying €15 with a 50% chance.

FIGURE 2.3 A typical ambiguity decision situation

To derive the indices of a subject's insensitivity and aversion, matching probabilities were elicited for all single $\{E_r, E_m, E_s\}$ and composite $\{E_{rm}, E_{ms}, E_{rs}\}$. For each single or composite event, a bisection was used to elicit its matching probability. For instance, for event E_r the subject first faced the decision situation depicted in Figure 2.3. If she chose option 1, in the next decision situation the winning probability in option 2 increased; otherwise, it decreased. For each event, subjects faced four decision situations, where option 1 stayed fixed and the winning probability in option 2 varied depending on the choices in the previous situation. Figure 2.A1 in Appendix 2.A shows how the probabilities for later decision situations and ultimately the event's matching probability were determined given subjects' choices. We will refer to the four decision situations for each event as a block.

The 24 decision situations for eliciting matching probabilities thus constituted 6 blocks. The blocks appeared in a random order, and between two consecutive blocks, a demographic question[6] was asked to refresh subjects' thinking mode. The demographic questions also appeared in a random order.

An example with explanation of the typical decision situation was presented to the subjects before they made their decisions. Subjects had to answer four questions checking their understanding correctly before they could proceed. Subjects could also click on a reminder button to view the description of the trust game again.

Following the matching probability decision situations, subjects in wave 2 made a decision as the trustee in the same trust game as before.[7] Figure 2.5 shows the trustee decision situation.

Subjects also answered non-incentivized WVS questions about their general trust attitudes. The three questions were: "Generally speaking, would you say that most people can be trusted or that you can't be too careful in dealing with people?"; "Would you say that most of the time, people try to be helpful, or that they are mostly just looking out for themselves?"; and "Do you think that most people would try to take advantage of you if they got the chance or would they try to be fair?". In each question, subjects could choose to agree or disagree with the statement. Answers indicating more trust were coded as 1, and 0 otherwise. The general trust measure was then taken as the average of the three responses.

[6] We asked five demographic questions: gender, alcohol consumption, general happiness question, nationality (Dutch or non-Dutch), and number of siblings.

[7] Subjects in wave 1 did not make the trustee decision. Subjects in wave 1 were part of a separate paper (Li et al. 2016), where the trust game studied in this paper was one of several treatments of that paper. The only difference in the design between wave 1 and wave 2 is that subjects in wave 1 did not need to make the trustee decision. Their trustee decisions are therefore treated as missing in the analysis of this paper.

FIGURE 2.5 Trust game: trustee decision situation

The following may be inside your envelope.

Recall that you are matched with one other participant. You can instruct the experimenters to give one of the following three options:

Option A: Pay **€15** to each of you
Option B: Pay you **€18**, pay your partner **€10**
Option C: Pay you **€22**, pay your partner **€8**

Your partner can instruct the experimenters to give you one of the following two numbered (1 and 2) options:

Option 1: Follow your instruction for payment
Option 2: Pay **€10** to each of you

The experimenters will follow your instruction only if your partner instructed to give you Option 1. If your partner instructed Option 2, then you and your partner will get €10 each, and your instruction will play no role.

Payment

After all subjects finished the experiment, they were called to the payment desk one by one. Each subject opened her envelope. If it was the trust game decision situation (either as the trustor or the trustee), her decision and her partner's choice would be used to determine her final payment. If the envelope contained a matching probability decision situation that she had encountered during the experiment, her partner's trustee decision determined her final payment had she chosen the ambiguous option 1. Otherwise, the winning probability of option 2 decided her payment.[8] It could also happen that the subject had not encountered the matching probability decision situation that came out of her envelope. In case this happened, we inferred the subject's choice in the new situation from her choice in a similar situation by dominance. For instance, suppose the subject had chosen option 1 in the decision situation in Figure 2.3, but a decision situation with a winning probability of 26% came out of her envelope. Because of the bisection procedure, she could not have encountered this situation during the experiment. We would then explain to the subject that, since she preferred the

[8] If, for instance, the winning probability of option 2 was 50%, then the subject threw two 10-sided dice, and any number below 50 (which had 50% chance of occurring) meant that the subject would be paid the prize.

ambiguous option 1 to an even better option 2 (with 50% winning chance), we inferred that in the decision situation where option 2 gives 26% winning chance, she would also prefer option 1. We would then implement option 1.

Discussion

The advantage of using Prince to implement the bisection procedure is that it enhances incentive compatibility. Under Prince, the decision situation that eventually mattered was pre-determined and did not depend on subjects' choices during the experiment, excluding the possibility to answer strategically so as to manipulate later stimuli. It was therefore always in the best interest of the subjects to reveal their true preferences.

2.4 Results

For each subject, we performed six monotonicity tests. By monotonicity, subjects' matching probabilities for a composite event should not be lower than those of the single events included in the composition. Therefore, two tests were performed for each composite event, resulting in six tests in total per subject. On average, the fail rate of these monotonicity checks was 7.5%. For the analysis reported below, we removed 20 subjects (11.0%) who violated monotonicity at least twice.

2.4.1 What determines decision to trust?

Table 2.1 shows summary statistics of all elicited variables.

TABLE 2.1 Summary statistics

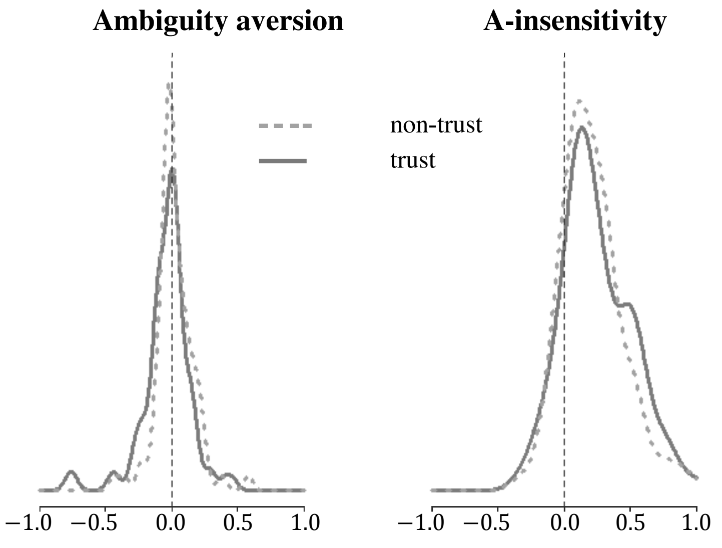
	Mean	Median	Std.dev.	Min	Max
Trusted	0.54	1	0.50	0	1
A.aversion	-0.01	0	0.17	-0.78	0.58
A.insensitivity	0.23	0.16	0.25	-0.32	1
p_r	0.31	0.32	0.21	0	1
p_m	0.30	0.33	0.16	0	0.96
p_s	0.41	0.33	0.24	0	1
$Pdiff$	-0.10	0	0.42	-1	0.96
General.trust	0.47	0.33	0.35	0	1
Male	0.56	1	0.50	0	1
Trustee	1.69	1	0.80	1	3
Weekly.drinks	4.18	2	5.15	0	30
Dutch	0.56	1	0.50	0	1
Happiness	7.01	7	1.67	0	10
Siblings	1.48	1	1.18	0	8

NOTES: Trusted = 1 if the trustor chooses the trusting option 1 and 0 otherwise; A.aversion is the ambiguity aversion index; p_r, p_m, and p_s are the a-neutral probabilities for the three events; $Pdiff = p_r - p_s$; General.trust is the average score in the WVS questions; Male=1 if the subject is male; Trustee = 1, 2, and 3 if trustee chooses option S, M, and R respectively, where a higher number corresponds to a more reciprocating option; Weekly.drinks is the average number of alcoholic beverage consumption; Dutch = 1 if the subject is Dutch, and 0 if not; Happiness is the subjective report to the question "Do you feel happy in general?", which can take values from 0 to 10; Sibling is the number of siblings.

Ambiguity-aversion and ambiguity-generated insensitivity

Figure 2.6 shows the density plots of the insensitivity and aversion indices grouped by subjects' decisions to trust. Trusting subjects are less ambiguity averse (p-value = 0.03; Wilcoxon one-sided test). There is no difference in insensitivity between trusting and non-trusting subjects (p-value = 0.44; Wilcoxon two-sided test).

FIGURE 2.6 Trust decision and ambiguity



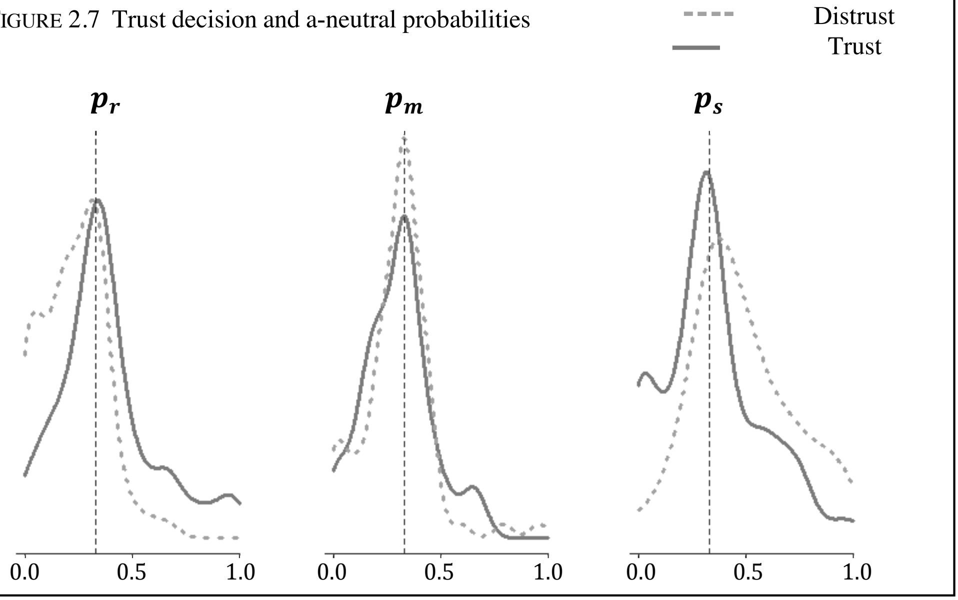
Beliefs

Figure 2.7 shows the density plot of a-neutral probabilities grouped by subjects' trust decisions. There was no difference between their beliefs about their trustee choosing option M (p-value = 0.9; Wilcoxon two-sided test). Trusting people believed their trustee to be more likely to choose the reciprocating option R (p-value < 0.01; Wilcoxon one-sided test), and, therefore, less likely to choose the selfish option S (p-value < 0.02; Wilcoxon one-sided test).

Table 2.2 presents marginal effects on subjects' decisions to trust in four logistic regression models (Trusted = 1). Model 1 includes as explanatory variables the two indices (ambiguity aversion and a-insensitivity) describing subjects' ambiguity attitudes; model 2 includes a variable that measures subjects' optimistic beliefs about others' trustworthiness ($Pdiff = p_r - p_s$); model 3 includes both ambiguity attitudes and beliefs; and model 4 adds five demographic variables to model 3.

The regression results are consistent with the non-parametric analysis and are robust across different model specifications. Subjects who are less ambiguity averse were more likely to trust others, but insensitivity did not matter. Subjects who believe their partners to be more reciprocating are also more likely to trust. Remarkable is that there was a systematic difference in the tendencies to trust between Dutch and non-Dutch subjects.

FIGURE 2.7 Trust decision and a-neutral probabilities



2.4.2 What are the World Values Survey questions measuring?

Table 2.3 reports four linear regression models on people's score in the WVS questions. Model 1 includes subjects' trust decisions in the trust game as an explanatory variable. Model 2 includes subjects' decisions as the trustee. Model 3 includes subjects' ambiguity aversion, a-insensitivity, and the differences in their p_r and p_s. Model 4 adds the five demographic variables to model 3.

Subjects' responses to the WVS questions are positively correlated with their decisions to trust, but have no relation with their own decisions as the trustee.[9] Subjects with more optimistic beliefs about their partners' trustworthiness score higher in the WVS questions, whereas their ambiguity attitudes play no role. These results are robust after controlling for demographic variables, showing that the trust survey questions do capture people's beliefs and are not distorted by ambiguity attitudes. There is a marginally significant gender difference with males scoring lower. Dutch subjects also score lower.

[9] The same holds if we include the trustee decisions as separate dummy variables in the regression.

TABLE 2.2 Regression: What determines decision to trust?

	Dependent variable: Decision to Trust			
	(1)	(2)	(3)	(4)
A.aversion	-0.529**		-0.592**	-0.592*
	(0.259)		(0.291)	(0.308)
A.insensitivity	0.126		0.178	0.164
	(0.168)		(0.184)	(0.191)
$Pdiff$		0.487***	0.519***	0.533***
		(0.115)	(0.122)	(0.127)
Male				0.146
				(0.090)
Weekly.drinks				-0.001
				(0.009)
Dutch				-0.193**
				(0.093)
Happiness				0.011
				(0.028)
Siblings				0.036
				(0.043)
Observations	162	161	161	161
Log Likelihood	-109.190	-100.208	-97.597	-94.253
Akaike Inf. Crit.	224.381	204.415	203.194	206.507

*p < 0.1; **p < 0.05; ***p < 0.01.

NOTES: We have 162 subjects left after removing those who failed monotonicity checks at least twice. Model 1 has 161 observations because for one subject, her a-neutral probabilities were non-identifiable since her matching probabilities for all events were the same.

2.4.3 How are beliefs formed?

People often form their beliefs about others based on their own types (Ross, Greene, and House 1977; Rubinstein and Salant 2016). Believing others to be similar to oneself may have contributed to the findings in Glaeser et al. (2000) that people's responses in the survey questions were correlated with their own trustee decisions. We find subjects' beliefs about their partners' trustworthiness to be strongly correlated with their own trustworthiness, although no correlation is found between their trust survey responses and their trustworthiness (Figure 2.8).

Subjects who choose option R believe their partners to be more likely to choose option R (p-value < 0.01; Jonckenheere test). Those who choose option S similarly believe others to be likely to do the same (choose S) (p-value < 0.01; Jonckenheere test), but not for option

M. In our sample, 21%, 27%, and 52% of the subjects chose option R, M, and S, respectively. These actual frequencies are closest to the median beliefs of those who chose option S as the trustee.

TABLE 2.3 Regression: What is the general trust survey measuring?

	Dependent variable: General Trust Survey			
	(1)	(2)	(3)	(4)
Trusted	0.110**			
	(0.055)			
Trustee		0.048		
		(0.039)		
A.aversion			-0.271	-0.199
			(0.173)	(0.174)
A.insensitivity			0.055	0.092
			(0.116)	(0.117)
$Pdiff$			0.149**	0.132**
			(0.065)	(0.065)
Male				-0.095*
				(0.056)
Weekly.drinks				0.009
				(0.006)
Dutch				-0.127**
				(0.061)
Happiness				0.014
				(0.017)
Siblings				0.026
				(0.024)
Constant	0.409***	0.560***	0.467***	0.410***
	(0.041)	(0.095)	(0.038)	(0.126)
Observations	161	125	160	160
R^2	0.024	0.012	0.050	0.104

*p<0.1; **p<0.05; ***p<0.01

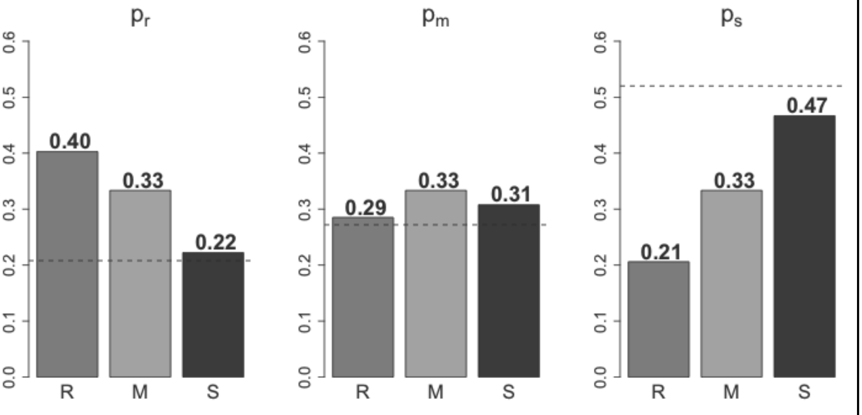
NOTES: The number of observations for model 2 is lower because 36 subjects in the first two sessions of the experiment did not make the trustee decision.

2.5 Discussion and related literature

Recently published theoretical studies that incorporate ambiguity into game theory include Angelopoulos and Koutsougeras (2015), Battigalli et al. (2015), Chakravarty and Kelsey (2016), De Marcoa and Romaniello (2015), Grant, Meneghel, and Tourky (2016), Kellner (2015), and Kelsey and le Roux (2016). In epistemic game theory (Perea 2012, 2014), beliefs themselves are Bayesian but unconventional updating methods are considered. Empirical studies incorporating ambiguity include Dai et al. (2015), Di Mauro and Finocchiaro Castro

(2011), Dominiak and Duersch (2015), Eichberger and Kelsey (2011), Eichberger, Kelsey, and Schipper (2008), Hsu et al. (2005), Kelsey and le Roux (2015, 2016), and Pulford and Colman (2007). Camerer and Karjalainen (1994) measured ambiguity aversion using an index of Dow and Werlang (1992). Baillon et al. (2016) showed that this index is a special case of their indices, also adopted in our paper. Whereas the method of Camerer and Karjalainen could only be used in the particular game they studied, ours can be used in all games. Ivanov (2011) measured ambiguity attitudes in a set of well-known (mis)coordination games such as the games of Stag Hunt, Chicken, and Matching Pennies. He controlled for beliefs by asking introspective questions. Our approach is entirely based on revealed preferences. The methods used in the previous studies measured only ambiguity aversion. We also measure insensitivity.

FIGURE 2.8 Belief about partner by own trustworthiness



NOTES: Each panel in Figure 2.8 presents the median a-neutral probabilities of an event (R, M, or S) split by subjects' own trustee decisions. The dashed horizontal line indicates the actual frequency.

Fehr (2009) reviewed existing literature on trust and argued that trust decisions are more than a special case of risk-taking. Economic primitives such as preferences and beliefs also play important roles. Our new ambiguity measurements show that this claim is correct: people's ambiguity aversion and beliefs about others matter for their trust decisions. The negative emotions aroused by the possibility of making themselves susceptible to betrayal are thus reinforced by not knowing the underlying chances.

Our finding that people with more optimistic beliefs (after correcting for ambiguity attitudes) about others' trustworthiness are more likely to trust is similar to what Sapienza, Toldra-Simats, and Zingales (2013) found. They used Berg et al.'s (1995) trust game where subjects could choose which part of their endowment to send to their trustee. To elicit subjects' beliefs about their trustees' trustworthiness, they used Selten's (1967) strategy method. They told subjects that their trustees would specify how much to return for every possible amount they receive. They then asked subjects to give a point estimate of the amount they expected their trustee to send back for every possible amount. To incentivize the expectations, they paid subjects for every accurate estimate that is not more than 10% off from the amount specified by the trustee. They found a positive correlation between subjects' expectation and the amount they sent. Our measure of belief is directly expressed in probabilities rather than indirectly in expectations and is directly revealed from choices. Further, we avoid income effects due to their multiple payments.

We shed new light on several open questions in the literature. One such question concerns whether introspective survey questions on trust, such as the ones included in the World Values Survey (WVS), are good measures of trust. A typical survey question asks "Generally speaking, would you say that most people can be trusted or that you cannot be too careful in dealing with people?". Findings on the correlations between answers to these questions and trust decisions measured in experiments using real incentives were mixed so far. Whereas Glaeser et al. (2000) and Lazzarini et al. (2005) found no correlations, Fehr et al. (2003) and Bellemare and Kröger (2007) found positive correlations. Because we can separate beliefs and ambiguity attitudes for the natural events relevant here, we can offer more refined insights. People's trust survey responses are positively correlated with their beliefs, and not with their ambiguity aversion or insensitivity. This finding confirms that the WVS questions are measuring trust in the commonly accepted sense of belief in others' trustworthiness, expressed for instance by Gambetta et al. (2000): "When we say we trust someone or that someone is trustworthy, we implicitly mean that the probability that he will perform an action that is beneficial or at least not detrimental to us is high enough for us to consider engaging in some form of cooperation with him."

We can also investigate how people formed their beliefs about others' trustworthiness. In the psychology literature, false consensus has been found, which describes people's tendency to expect others to be close to themselves in characteristics, preferences, and so on (Ross, Greene, and House 1977). For instance, people who are happy themselves would expect a larger proportion of the population to be happy than would unhappy people. Although the name of this phenomenon suggests that it is a bias in judgment formation, later studies showed that it could be the result of rational Bayesian updating using one's own type as a signal (Dawes 1990; Prelec 2004). Similar to Rubinstein and Salant (2016), we find support for the self-similarity reasoning in our game theoretical setting: people's belief about others' trustworthiness is correlated with their own trustworthiness. Thus the beliefs of the most prevalent type—the non-trustworthy one—are close to the actual distribution of trustworthiness in our sample, and own type serves as a useful signal here too.

Another contribution of the current paper is that we show how the prior-incentive system (Prince), which was initially designed for individual decision making, can be adapted to game experiments. The prior-incentive system enhances the incentive compatibility of the random incentive system. We have shown how it can be implemented in strategic situations, where subjects have to be matched with partners or opponents and strategic confounds must be avoided.

2.6 Conclusion

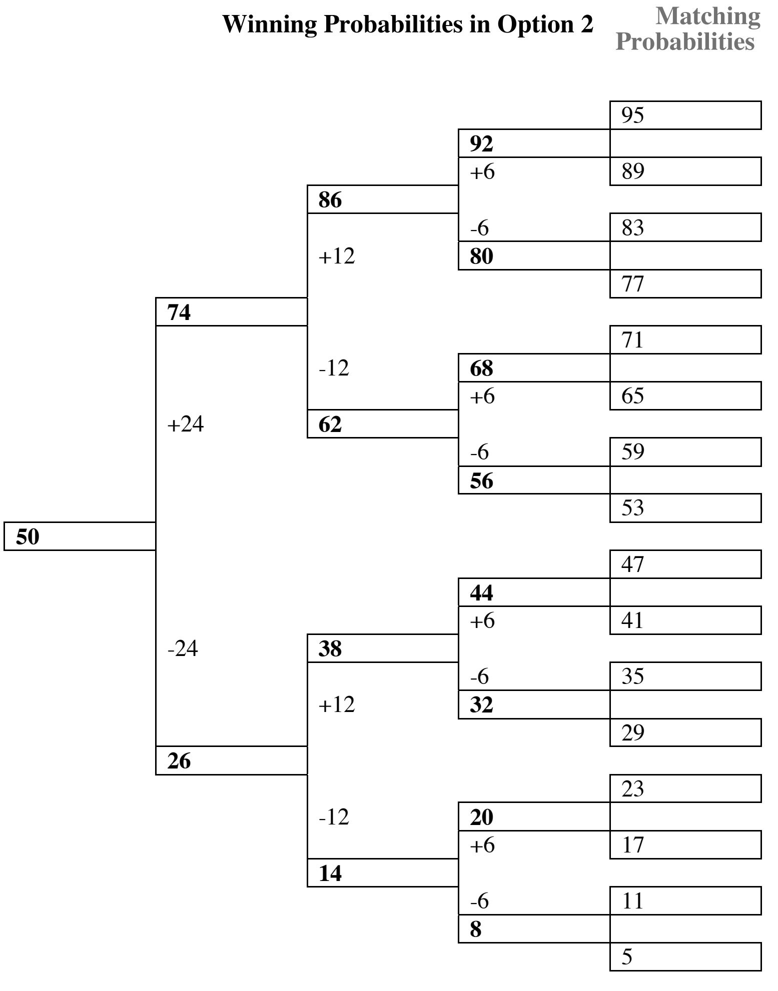
Decisions to trust are almost always decisions under ambiguity, because we do not know the probabilities of others being trustworthy. Studies have so far focused on relations between risk attitudes and trust, finding little relation (Eckel and Wilson 2004; Houser et al. 2010). No method was known to measure ambiguity attitudes in natural situations such as trust decisions. Ambiguity attitudes and beliefs remained as uncontrolled confounds.

We used Baillon et al.'s (2016) method that can measure ambiguity attitudes in natural situations. It therefore opens up the possibility to analyze (and correct for) ambiguity attitudes in experimental games. This holds in particular for belief measurements (e.g. of another person being trustworthy). Belief measurements are widely used in experimental economics, but always under the assumption of ambiguity neutrality so that ambiguity attitudes confound them. We have extracted beliefs while controlling for ambiguity.

The motivational ambiguity aversion contributes to people deciding not to trust others. The cognitive insensitivity (perception and understanding) does not impact such decisions. We confirm that people who have more optimistic beliefs about others' trustworthiness are more likely to trust others. We shed new light on several unsettled issues in the literature. First, based on revealed preference data we show that the survey trust questions do capture people's beliefs about others' trustworthiness. Moreover, people's beliefs about others are positively correlated with their own trustworthiness. Hence, own type serves as a useful signal about others.

2.A Appendix

FIGURE 2.A1 Determination of probabilities in the bisection method



NOTES: For each event, the winning probability of the first decision situation is always 50%. At each node, if the subject chooses option 1 (2), the probability on the upper (lower) branch is used as the winning probability in option 2 in the next decision situation, while option 1 remains the same. The last column is the matching probability recorded depending on subjects' choices in the previous four decision situations.

Chapter 3

Social and Strategic Ambiguity: Betrayal Aversion Revisited

with CHEN LI AND PETER P. WAKKER

3.1 Introduction

Ellsberg (1961) showed that people treat ambiguity—uncertainty with no objective probabilities available—in a fundamentally different way than risk where objective probabilities of uncertain events are known. Ambiguity attitudes describe these differences. A big field of application for ambiguity theory is game theory, where the uncertainty is strategic. That is, uncertainty concerns the strategy choice of an opponent who interacts with the decision maker, and who may have common or opposite interests. Game theory has traditionally made the idealized assumption that all strategic uncertainties can be modeled as risk. Probabilities of strategy choices of opponents are however rarely known in reality, and therefore most games are better modeled as decision under ambiguity.

Whereas many theoretical papers applied ambiguity theories to game theory there have as yet only been few empirical studies doing so (e.g., Dominiak and Duersch 2015, Chark and Chew 2015, Eichberger and Kelsey 2011, Ivanov 2011, and Kelsey and le Roux 2015). In particular, the popular measurements of subjective beliefs in games are invariably done under the assumption of ambiguity neutrality (Trautmann and van de Kuilen 2015b footnote 16). The reason is that no method was known to measure ambiguity attitudes except for in some artificially constructed situations (Ellsberg urns and experimenter-specified probability intervals) where a control for belief based on symmetry was readily available. Baillon et al. (2016) introduced a method for measuring ambiguity attitudes in general situations of individual decisions that need no artificial symmetry, and Li, Turmunkh, and Wakker (2016) extended this method to game theory. The latter paper applied their method to trust games, disentangling and controling for different components that may confound the measurement of trust. We use Li et al.'s (2016) method and investigate how ambiguity attitudes differ in games against interested others versus games against "neutral nature." We identify a new non-strategic component underlying all strategic ambiguities, called social ambiguity, discussed in §3.2.

We use our new measurement and controls to resolve some unsettled questions about Bohnet and Zeckhauser's (2004; B1) and Bohnet et al.'s (2008; B2) finding of betrayal aversion. They developed a clever design using minimally acceptable probabilities (MAPs) to measure betrayal aversion while correcting for risk attitude. Their theoretical analyses were based on the classical models with rational ambiguity neutrality that were commonly used to analyze games and incentive compatibility. We provide a re-analysis using modern ambiguity theories and show that ambiguity attitudes rather than betrayal aversion can explain their findings. We thus confirm the findings of Fetchenhauer and Dunning (2012). They used stimuli that did avoid ambiguity and genuinely involved only risk, and found no betrayal aversion. Our contribution to their study is to analyze trust decisions under ambiguity, which is more realistic. In particular, we can investigate how betrayal aversion interacts with ambiguity. We conclude that there is no systematic betrayal aversion. The positive emotions about gaining due to trustworthiness apparently offset the negative emotions about losing due to betrayal. While betrayal ambiguity does not affect behavior motivationally (no aversion to betrayal), we find that it does affect behavior cognitively: people are insensitive to different likelihoods of others' trustworthiness, and apparently strategic ambiguity is harder to process than nonstrategic ambiguity. Our control for social ambiguity shows that the latter effect is specifically induced by strategic ambiguity and not by general social ambiguity.

This paper proceeds as follows. Section 3.2 discusses social and strategic ambiguity in games. Section 3.3 describes how we measure ambiguity attitudes. Section 4 describes our experiment. The data and findings are reported in Section 3.5, and discussed in Section 3.6. Section 3.7 concludes. Although our theoretical re-analysis of B1 and B2 was an important motivation for us to undertake this study, we present it in Appendix 3.A so as to have the main text accessible to experimentally oriented readers.

3.2 Strategic and social ambiguity

The aforementioned idealization—reduction of ambiguity to risk—commonly adopted in game theory can be achieved by assuming rationality and common knowledge. Luce and Raiffa (1957, p. 306) describe how in a game between rational and knowledgeable actors the ambiguity about the opponent's intended act reduces to risk:

"One *modus operandi* for the decision maker is to generate an *a priori* probability distribution over the states (pure strategies) of his adversary by taking into account both the strategic aspects of the game and what 'psychological' information is known about his adversary, and to choose an act which is best against this *a priori* distribution. To determine such a subjective *a priori* distribution, the decision maker might imagine a series of simple hypothetical side bets whose payoffs depend upon the strategy his adversary employs. … until there exists an equilibrium in the decision maker's mind. … If in a given situation the theory is clear cut and if a decision maker knows that his adversary will comply with the theory, then in a sense, the theory defines the decision maker's choice of an *a priori* distribution for his adversary."

In applications, probabilities of others' strategy choices are virtually never available. Even when preferences are known, the decision maker may not be able to navigate effortlessly through belief hierarchies and arrive at the "mind equilibrium" envisioned by Luce and Raiffa (1957).[10] And if she is capable of doing this, she would not be confident that her opponent is. If the opponent's mind works differently then she has no reason to believe that the deduced probabilities are objective, and they may not be shared by the opponent. In reality, decisions in games are made on the basis of subjective beliefs, and are made under ambiguity rather than risk. Accordingly, measuring and studying the actors' ambiguity attitudes is important for understanding and predicting their decisions in a game.

Ambiguity attitudes have been studied much in the context of individual decision making, where the uncertainty facing the decision maker concerns the act of a neutral "nature." This paper investigates how ambiguity attitudes are different when people play against interested others than when they play against nature. In strategic interactions, the decision maker must reckon with the uncertain interests of others, and also with the others' beliefs about her interests, their beliefs about her beliefs about their interests, and so forth. Nature by contrast has no interests, and does not care about the interests of the decision

[10] This topic is central in epistemic game theory (Perea 2012, 2014).

maker. Yet, differently from an uninterested player, nature chooses some of its acts with higher likelihoods than others. Therefore, strategic complications make ambiguity in games against interested others fundamentally different from ambiguity faced in a game against nature. For further discussions of the differences between natural and strategic uncertainty, see Aumann and Drèze (2009), Gilboa and Schmeidler (2003), Harsanyi (1982), Schneeweiss (1973), Sugden (1991 §XI, p. 782 bottom), and von Neumann and Morgenstern (1944 p. 11, p. 99).

A difference between strategic and nature uncertainty was suggested before by the finding that risk attitudes (measured from choice over "nature" lotteries) could not explain people's decisions in games against others (Eckel and Wilson 2004, Houser et al. 2010). Papers on the other vs. nature divide as yet sought to study this issue within the risk domain; that is, to compare willingness to take risk in a situation where the probabilities of monetary outcomes were generated by choices made by opponents to a situation where the same probabilities of the same outcomes were generated by a chance device. Any difference was then attributed to the additional emotional consequences that people experienced when outcomes resulted from interactions with others. Thus, B1 and B2 found that people required a higher probability of winning from a "human" lottery than from a "nature" lottery (see also B2). The authors interpreted their stimuli as risky and attributed their finding to the additional emotional costs when a bad outcome resulted from another's act (betrayal) rather than from forces of nature.

We believe that ambiguity is a more natural domain for studying the other vs. nature divide in people's behavior under uncertainty. B1 and B2 used MAPs to avoid ambiguity (B2 footnote 3) but according to modern theories, ambiguity still was present (Appendix 3.A), as were attitudes towards complexity and dynamic optimization under ambiguity. These other factors instead of betrayal aversion can explain their findings. Fetchenhauer and Dunning's (2012) stimuli present the other person's options as having been drawn from a known distribution. Thus they succeeded in genuinely avoiding ambiguity. But their scenario is not realistic and will rarely occur in social decisions. We therefore use a design where trust decisions are made as in a typical situation, with no artificial objective probabilities provided.

B1 and B2 interpreted the effects that they found, broadly, as a difference between social and nature risk, but also specifically as (strategic) betrayal aversion. We investigate a potentially important second, non-strategic, reason that makes uncertainty in games against others distinct from that in games against nature. It is that people treat acts of humans, also when strategic complications are absent, in a fundamentally distinct way from acts of nature, which are free from human agency and free will. Greek philosophy emphasized the distinction between law of nature ("physis") and law by humans ("nomos"). What was "by nature" was considered unquestionably right and hence universally binding, whereas laws made by humans could be deemed right by some but wrong by others. In modern legal traditions the principle of "act of God" or "superior force" is used to determine absence of human agency. The basic notion that forces of nature are beyond judgment extends to our everyday understanding of causes and consequences, in which presence of human agency inextricably engenders various emotional judgments of the generosity, decency, or propriety of others' acts. Phenomena ascribed to the strategic aspects of the game may in reality have been driven by this general social ambiguity.

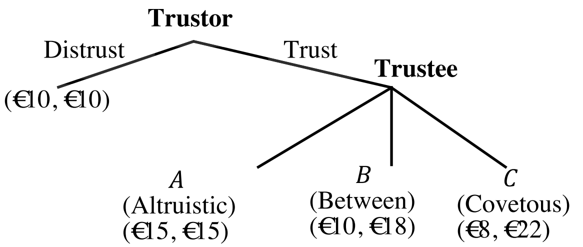
To parse out social versus strategic ambiguity in games and, hence, to identify what is truly caused by the strategic aspects of the interactions, our experiment includes the following three treatments: nature ambiguity, social ambiguity but with no (strategic) betrayal or special other emotions involved, and, finally a social treatment with also betrayal involved. This allows us to separate strategic attitudes, more specifically, betrayal aversion, from purely social attitudes and, then, from nature attitudes.

We use indices of Baillon et al. (2016), explained in §3.3, that capture aversion for particular kinds of uncertainty, such as about betrayal, through event weighting functions. This builds on the source preference concept first coined by Heath and Tversky (1991) for a preference of basketball fans for basketball uncertainty, and formalized and axiomatized by Tversky and Wakker (1995). The literature on betrayal aversion usually tried to model such source preference through utilities of outcomes. Although at first sight this modeling may seem plausible, and hence has often been suggested informally for ambiguity (Smith 1969), there are some difficulties. One is that there is no clear way to capture such modeling in an overall decision theory for ambiguity satisfying basic conditions such as transitivity, monotonicity, continuity, and consistency for sure outcomes across different underlying sources of uncertainty. Source preference through weighting functions is part of prospect theory, which satisfies all aforementioned conditions (Tversky and Wakker 1995). A second difficulty is that utilities of outcomes cannot capture an important insensitivity component. For instance, Fetchenhauer and Dunning (2012) found that for a small chance of winning, subjects were more willing to take the betrayal risk rather than the nature risk, but when the chance of winning was large there was no difference in willingness to take risk. Such a dependency on likelihood cannot be captured by outcome functions. In particular, utility cannot capture the difference between social ambiguity with and without ambiguity, which concerns insensitivity.

3.3 Measuring ambiguity attitudes

Consider the two-player trust game in Figure 3.1. Throughout, $(€\alpha, €\beta)$ denotes an outcome where the trustor receives $€\alpha$ and the trustee $€\beta$. We will only consider gains $\alpha \geq 0, \beta \geq 0$. If the trustor in this game chooses the act Distrust, it may be because she expects the trustee to choose the covetous act C. It may also be the case that her best guess is that the trustee reciprocates her trust with the altruistic act A, but she is not confident about her best guess and dislikes the ambiguity enough to go for the certain payoff $(€10, €10)$ of Distrusting. If the trustor instead chooses the act Trust, it is similarly unclear to what extent her ambiguity attitude affected this decision.

FIGURE 3.1 Trust game

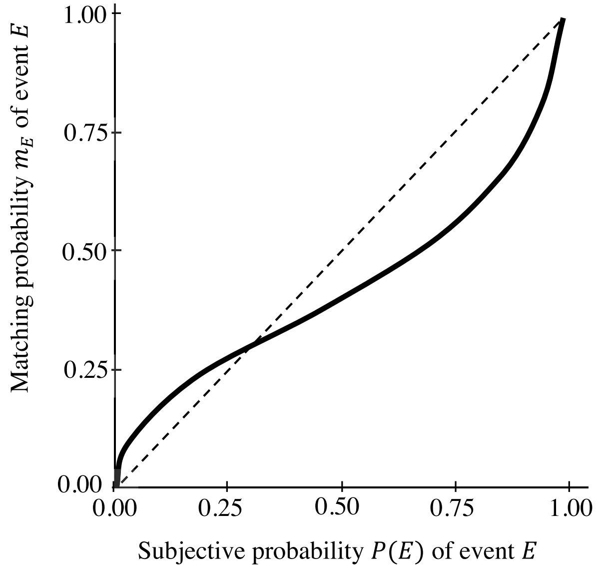


We now describe how appropriately constructed side bets can be used to measure the ambiguity attitude of the trustor, with the help of matching probabilities. A decision maker's *matching probability* of event E is the probability m_E that makes the decision maker indifferent between the ambiguous lottery (E, x) offering prize x if event E occurs and nothing otherwise, and the risky lottery (m_E, x) offering the (same) prize x with probability m_E and nothing otherwise. That is, the decision maker's matching probability m_E of event E is the probability m_E such that $(E, x) \sim (m_E, x)$. Under Savage's (1952) subjective expected utility (SEU)—i.e., ambiguity neutrality—the ambiguous lottery (E, x) is evaluated by its SEU, $P(E)U(x) + (1 - P(E))U(0)$, with $P(E)$ the decision maker's subjective probability of event E and U her utility function. The risky lottery (m_E, x) is evaluated by $m_E U(x) + (1 - m_E)U(0)$. The matching probability m_E of an ambiguity-neutral decision maker then measures her subjective probability of event E: $m_E = P(E)$. Empirically we usually find $m_E \neq P(E)$, because most people are not ambiguity neutral. The size and sign of the deviation reflect ambiguity attitude, with for instance $m_E < P(E)$ under ambiguity aversion.

Figure 3.2 shows a graph of the commonly observed deviations of matching probabilities from subjective probabilities. The dotted line in the figure represents the matching probabilities of an ambiguity-neutral decision maker, who treats unknown (subjective) probabilities as if they were known. Empirical studies have found that people commonly behave as represented by the solid line (Trautmann and van de Kuilen 2015a, Wakker 2010 §10.4.2), displaying ambiguity-aversion for likely events and ambiguity-seeking for unlikely events. The commonly observed ambiguity attitude described by the solid line captures the well-known Ellsberg paradox. For instance, in Ellsberg's two-color urn problem regardless of the winning color people prefer to gamble on the known urn with equal numbers of green and red balls over the unknown urn. Such preferences can be accommodated by any matching probability function (including the one depicted in Figure 3.2) with $m_E < 0.5$ for event E and with subjective probability $P(E) = 0.5$.[11]

[11] Let G and R denote the events that a green ball and a red ball is drawn, respectively, from the unknown urn. Chew and Sagi (2008) showed that the events can have subjective probabilities $P(G) = P(R) = 0.5$, but still the decision maker's matching probabilities m_G and m_R of the two events can both be less than 0.5—i.e., $(G, x) \sim (m_G, x)$, $(R, x) \sim (m_R, x)$, and m_G, $m_G < 0.5$—then $(G, x) \prec (0.5, x)$ and $(R, x) \prec (0.5, x)$. In particular, contrary to what has been believed long time, the Ellsberg paradox can be reconciled with subjective probabilities.

FIGURE 3.2 Commonly observed matching probabilities



Thus, deviations of matching probabilities from the decision maker's subjective probabilities assigned to the corresponding events (the 45-degree dotted line) describe the decision maker's ambiguity attitude. The relative elevation of the matching probability function captures the extent to which the decision maker likes or dislikes ambiguity (motivational component), and the relative flatness in the middle captures the (in)sensitivity of the decision maker, who insufficiently discriminates different levels of likelihood of ambiguous events (cognitive component). The motivational component reflects ambiguity aversion/seeking. We call the cognitive component *ambiguity-generated likelihood insensitivity*, or *a-insensitivity* for short.[12]

For measuring ambiguity attitudes in a game against others a major challenge lies in the difficulty of controlling for the decision maker's subjective probabilities. Baillon et al. (2016) proposed a method that overcomes this challenge. They showed that matching

[12] Gonzalez and Wu (1999) provide a clear discussion of these psychological interpretations of elevation and curvature for risk attitudes. Their concepts are naturally extended to other weighting functions (Wakker 2010 §10.4), being matching probabilities in our case.

probabilities of three events, constituting a partition of the state space, and their unions are sufficient for quantifying the decision maker's ambiguity aversion and a-insensitivity. In the trust game of Figure 3.1, it is therefore sufficient to elicit the trustor's matching probabilities m_A, m_B, and m_C of the three single events A, B, and C corresponding to the three possible acts of the trustee, together with the matching probabilities m_{AB}, m_{AC}, and m_{BC} of the three composite events (A or B), (A or C), and (B or C). For a detailed discussion of how ambiguity attitudes can be measured in strategic games, see Li et al. (2016).

Baillon et al. (2016) proposed the following two indices to quantify ambiguity aversion (b) and a-insensitivity (a):

$b = 1 - (m_s + m_c)$,

$a = 1 - 3(m_c - m_s)$.(3.1)

Here $m_s = (m_A + m_B + m_C)/3$ denotes the average single-event matching probability, and $m_c = (m_{AB} + m_{AC} + m_{BC})/3$ the average composite-event matching probability. The ambiguity aversion index b captures the general elevation of the decision maker's matching probabilities, using the sum $m_s + m_c$ which, under ambiguity-neutrality, equals 1 so that $b = 0$. The a-insensitivity index a captures the flatness of the matching probability function in the middle region, using the difference $m_c - m_s$ which, under ambiguity-neutrality, equals 1/3 so that $a = 0$. Positive values of b and a describe ambiguity-aversion and a-insensitivity. The larger the values, the more the decision maker deviates from ambiguity-neutrality.

3.4 Experiment

Subjects

N = 248 students from Erasmus University of Rotterdam took part in the experiment. The experiment was computerized and involved three treatments—betrayal, social, and nature—corresponding to the three different types of ambiguity that subjects faced.[13] Each subject was randomly assigned to one and only one treatment. $N_1 = 88$, $N_2 = 80$, and $N_3 = 80$ were assigned to betrayal, social, and nature ambiguity, respectively.

[13] Experimental instructions are in the online appendix of the paper.

Treatments

In each treatment, a subject faced a triple of mutually exclusive and exhaustive events A, B, and C, only one of which was true. Subjects did not learn the true event until the end of the experiment.

In the nature ambiguity treatment, ambiguity was generated by the hidden marking of a card drawn by each subject at the start of the experiment. The card was drawn from a deck of four cards, which could be marked with the letter A, B, or C. It was ambiguous to the subject how many of the cards in the deck were marked with each of the three letters.

Subjects in the other two treatments faced social ambiguity. They were randomly assigned an anonymous partner (a fellow subject) at the start of the experiment. The events A, B, or C described the choice made by the partner. Same as in nature ambiguity, the ambiguity facing the subjects was presented to them at the start of the experiment, and resolved at the end. Subjects assigned to social ambiguity (without betrayal) were told that the assigned partner would choose one of three snacks labeled A, B, and C. The partner's choice involved no betrayal, only the consideration of which of the three snacks would be most desirable to her. In the third treatment subjects faced betrayal ambiguity. Here, subjects were presented with the description of the game and were told that the partner would choose one of the three allocations labeled A, B, and C in the trust game (Figure 3.1).[14]

Procedure

Figure 3.3 presents the procedure for the three treatments. In step 1, subjects were given general instructions, common for all three treatments. In step 2, the treatment-specific ambiguity was presented. In step 3, six matching probabilities were elicited from each subject, with the contingency events of the ambiguous lotteries varying in accordance with the subject's ambiguity treatment. The matching probabilities in this procedure will be explained later. Finally, in step 4 subjects in the social ambiguity treatments made their ambiguity-generating choices as a partner to another subject. The partner pairing in the two social ambiguity treatments was such that each subject also made the ambiguity-generating choice as a partner to another subject. The choice as a partner—of one snack or one allocation

[14] We did not use the terms trust/distrust and altruistic/between/covetous for the subjects in the experiment, but only the neutral letters as labels.

in the role of a trustee—to another subject was made at the end of the experiment, in step 4.[15]

FIGURE 3.3 Summary of experimental procedure

Ambiguity Treatment	Nature	Social	Betrayal
Step 1.	General instructions		
Step 2. Presentation of ambiguity	Subjects drew a card from an ambiguous deck of four cards. It was ambiguous whether the card drawn was marked with letter A, B, or C.	Subjects were informed that they were matched with a partner, and told that the partner would choose one desired snack out of three: salty snack, dried fruit snack, and a chocolate bar, labeled A, B, or C. It was ambiguous whether the partner would choose snack A, B, or C.	Subjects were informed that they were matched with a partner, and told that the partner would choose one of three allocations: (€15, €15), (€10, €18), or (€8, €22), labeled A, B, or C. Subjects could choose a safe allocation of (€10, €10) or let the partner's choice determine the allocation. It was ambiguous whether the partner would choose allocation A, B, or C.
Step 3. Measurement of ambiguity attitude	Elicitation of matching probabilities of the six events A, B, C, (A or B), (A or C), and (B or C).		
Step 4. Ambiguity-generating choice as a partner to another subject		Subjects chose one desired snack out of three: salty snack, dried fruit snack, and a chocolate bar, labeled A, B, or C.	Subjects chose one of three allocations: (€15, €15), (€10, €18), or (€8, €22), labeled A, B, or C.

[15] Half of the subjects in the social ambiguity treatment were paired with subjects in the betrayal ambiguity treatment, so that in step 4 they chose not a snack but an allocation in the role of a trustee. Similarly, in these sessions subjects facing betrayal ambiguity chose in step 4 a snack and not a trust game allocation as a trustee. The ambiguity-generating choice as a partner to another subject is not relevant for our analysis, and was always made at the end of the experiment, so that subjects' responses to the matching probability elicitations (focus of our analysis) were not affected by the ambiguity-generating choices that they had to make as partners to the other subjects.

Matching probabilities and incentives:

Matching probabilities were elicited as follows. In all three treatments subjects faced a triple $\{A, B, C\}$ of mutually exclusive and exhaustive ambiguous events. From each subject we elicited the following six indifferences:

$$(A, €15) \sim (m_A, €15),$$
$$(B, €15) \sim (m_B, €15),$$
$$(C, €15) \sim (m_C, €15),$$
$$(A \text{ or } B, €15) \sim (m_{AB}, €15),$$
$$(A \text{ or } C, €15) \sim (m_{AC}, €15),$$
$$(B \text{ or } C, €15) \sim (m_{BC}, €15).$$

All indifferences were elicited by means of bisection; that is, subjects were presented with a binary choice between an ambiguous lottery (E, x) and a series of risky lotteries (m, x), whose probability m of winning the prize x was adjusted with each subsequent step of the bisection until the matching probability m was obtained such that $(m, x) \sim (E, x)$. We used a four-step bisection, and the binary choice in the first step used the risky lottery $(0.50, €15)$. If a subject preferred the ambiguous lottery, then in the second step the risky lottery was made more attractive. If the subject instead preferred the risky lottery, then in the second step she faced a choice between the ambiguous lottery and a less attractive risky lottery. Thus, the winning chance of the risky lottery was adjusted in each subsequent step, with the size of the adjustment shrinking at each subsequent step—from +/–0.24 to +/–0.12 and +/–0.06. Following the fourth choice, the matching probability could be inferred within +/–0.03 bounds.

Bisection has the practical advantage of being an efficient tool for measuring indifferences that is also easy for subjects to understand. However, the chained elicitation of bisection, if not carefully administered, can be problematic for real incentives. Specifically, in our elicitation of matching probabilities, which were adjusted according to previous choices of participants, one may be concerned about it being advantageous for participants not to answer according to their true preferences but instead to seek to improve the stimuli (risky prospects) that will occur in future choices. To cancel the possibility of such strategic behavior, we adopted in our experimental design the prior-incentive (Prince) system proposed by Johnson et al. (2014).

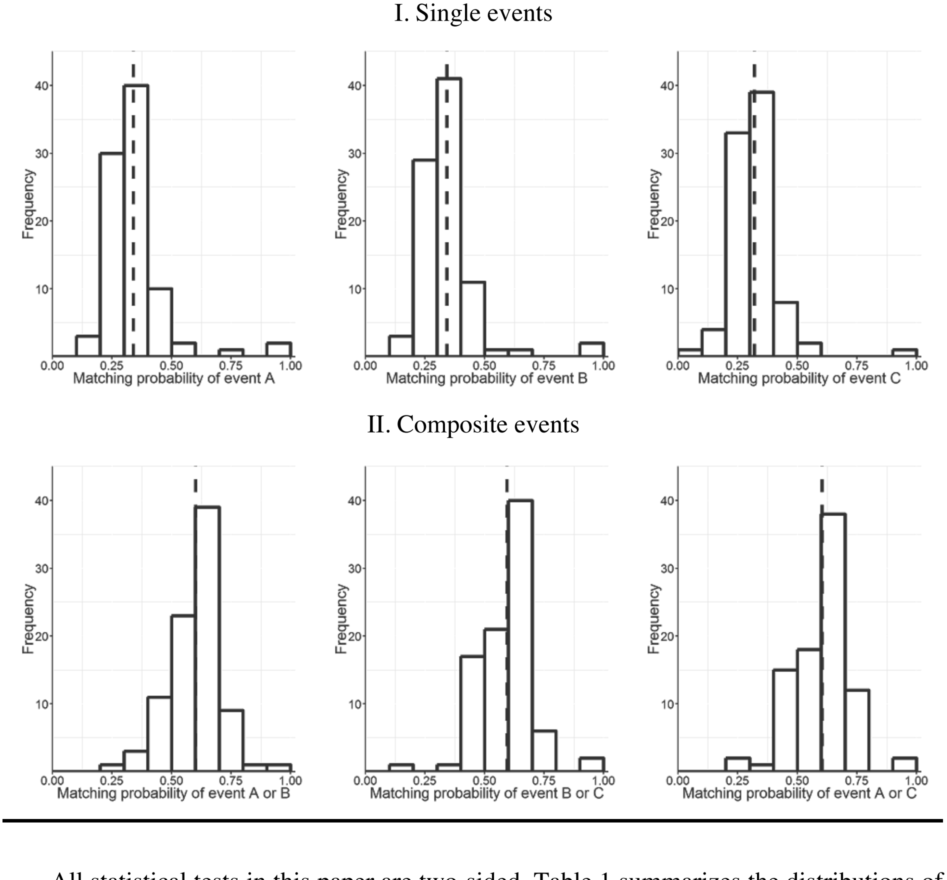
Prince combines the spirit of the random-incentive system, in which only one of the experimental choice questions is implemented for real, with the added benefit of preventing subjects from answering strategically in adaptive experiments (where the sequence of stimuli presented to participants is path-dependent). The key element of Prince is that the choice question implemented for real is randomly selected *prior* to the experiment. In our experiment, the real choice question was provided to each participant at the start of the experiment in a sealed envelope, which the participant randomly selected from a box containing the envelopes. Thus, participants knew that they could not influence the choice question that was to be played for real.

3.5 Results

Our main data are presented in Figures 3.4-3.6. Figure 3.4 shows the distributions of the six matching probabilities for nature ambiguity. Figures 3.5 and 3.6 show the matching probabilities for social and betrayal ambiguities. Matching probabilities express subjects' willingness to bet on ambiguous events; they capture both subjective probabilities and ambiguity attitudes of the decision makers. For nature ambiguity (Figure 3.4), although subjects' matching probabilities are non-additive (due to ambiguity attitudes), they display the sort of symmetry expected in an ambiguous situation where the decision maker has no reason to believe that one event is any more likely to occur than another event. In this case, subjects had no reason to believe that the card drawn was any more likely to be marked with letter A than with letter B or letter C. The matching probabilities in Figure 3.4 show that subjects indeed displayed equal willingness to bet, reflecting symmetric beliefs with regard to all single and all composite events: the average matching probability was about 0.33 for each of the single events, and 0.60 for each of the composite events.

For social and betrayal ambiguities the symmetric beliefs were no longer plausible. Figures 5 and 6 confirm this. In both treatments, subjects tended to display a higher willingness to bet on one particular event, suggesting that they believed one event to be more likely to occur than any of the other events. For social ambiguity (Figure 3.5) subjects were most willing to bet on event C (that the partner would choose the chocolate bar over the other two snacks). For betrayal ambiguity subjects were most willing bet on event C (that the partner would choose the selfish allocation C). It is interesting to observe that the asymmetries exhibited by the subjects' matching probabilities were "reasonable" in the sense that the average matching probabilities (dashed line) were for all events close to the actual choice fractions of the partners (solid line). Finally, we also observe that the matching probabilities for social and betrayal ambiguities were more dispersed than the matching probabilities for nature ambiguity. Clearly there was more heterogeneity in beliefs in the situations of social ambiguity than in the case of nature-generated ambiguity.

FIGURE 3.4 Matching probabilities for nature ambiguity



All statistical tests in this paper are two-sided. Table 1 summarizes the distributions of subjects' ambiguity attitudes. The top panel shows the ambiguity aversion index. For nature ambiguity we find ambiguity aversion (b index was positive), which is consistent with the preceding literature that studied ambiguity attitudes using Ellsberg urns. Toward social and betrayal ambiguities, however, there was no aversion. Further, the bottom panel of the table shows that the more important component driving the deviation from ambiguity-neutrality was cognitive, rather than motivational. A-insensitivity was significant (a index was positive) for all ambiguities.

FIGURE 3.5 Matching probabilities for social ambiguity

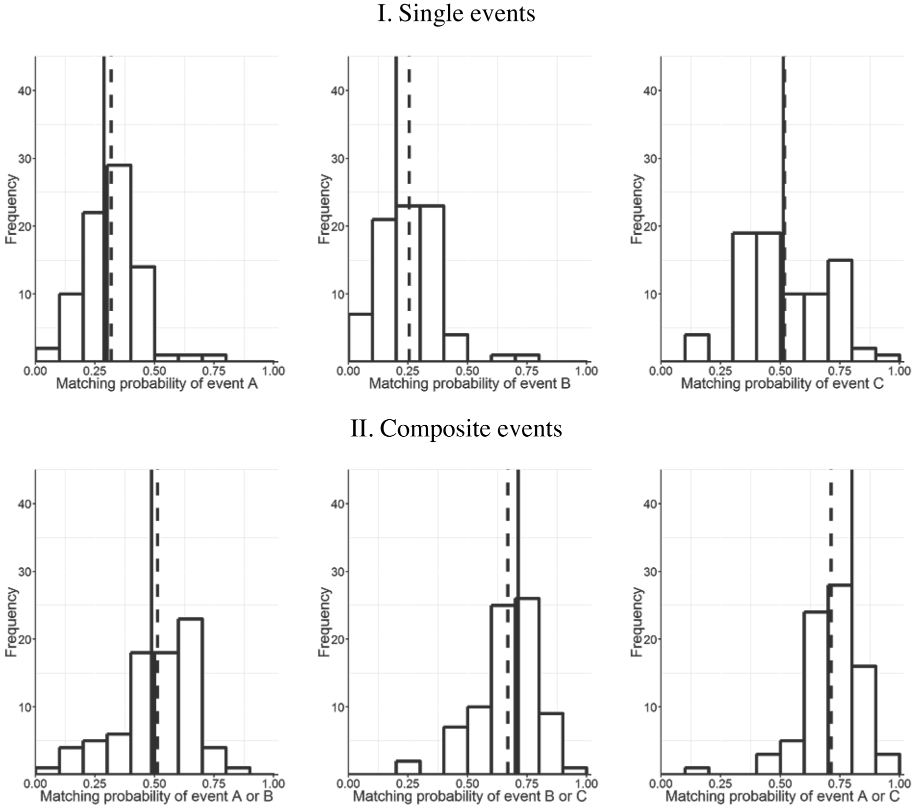
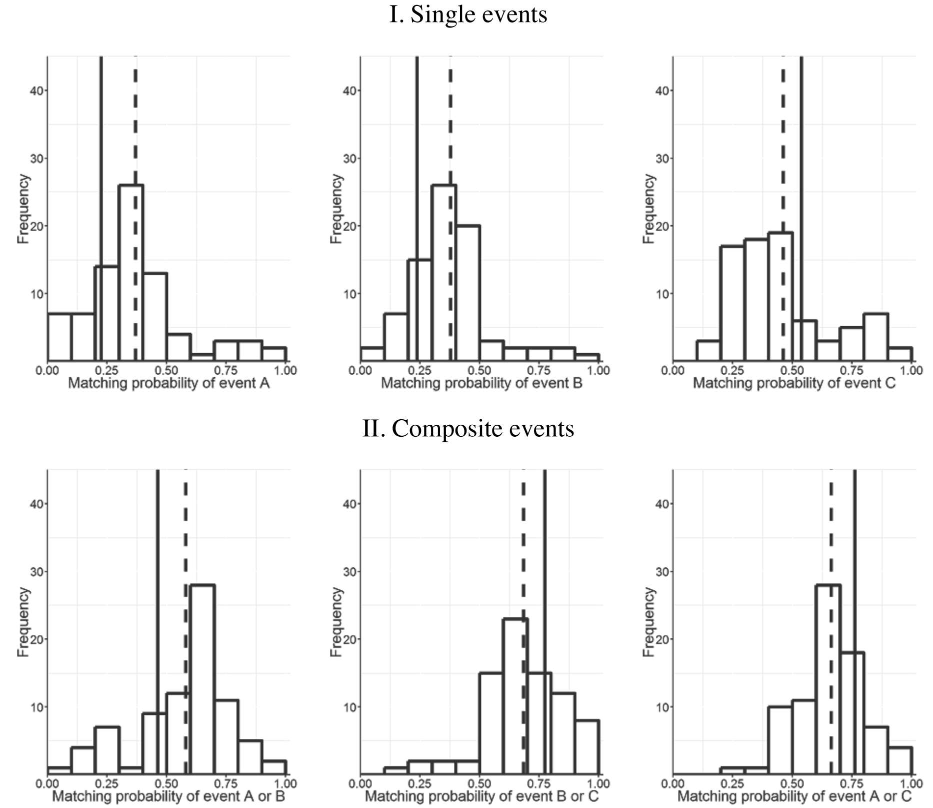


Figure 3.7 compares the ambiguity attitudes of nature vs. social and betrayal. The left panel shows the differences in ambiguity aversion. Subjects were less ambiguity averse when the ambiguity concerned a partner's choice in a trust game (toward betrayal ambiguity), than when it concerned the marking of a card drawn from an unknown deck (p-value < 0.001; Mann-Whitney U test). Comparison of the nature ambiguity with social ambiguity shows a similar difference, with subjects displaying less ambiguity aversion toward social ambiguity than toward nature ambiguity (p-value = 0.02). When we compare

social ambiguity with betrayal ambiguity, we find no difference in ambiguity aversion (p-value = 0.16). Our conclusion is that there was no betrayal aversion (no systematic preference for nature ambiguity over betrayal ambiguity). In fact, our results show the opposite preference: preference for social and betrayal ambiguities over nature ambiguity. The equal preference for social and betrayal ambiguities suggest that the motivational difference was driven by general social ambiguity rather than by strategic ambiguity.

FIGURE 3.6 Matching probabilities for betrayal ambiguity



The right panel of Figure 3.7 shows the differences in the cognitive component of ambiguity attitude, a-insensitivity. Subjects displayed a-insensitivity under all ambiguities (Table 1). They were more insensitive for betrayal than for social (p-value = 0.03) or nature (p-value = 0.02) ambiguity. Social ambiguity played no role in the cognitive divide, as there was no difference in insensitivity under nature and social ambiguities (p-value = 0.70). We conclude that whereas social ambiguity affects the motivational divide in other vs. nature ambiguity attitude, the presence of strategic complications in games against others leads to more a-insensitivity. Furthermore, stronger insensitivity under betrayal ambiguity implies that people prefer betrayal ambiguity over nature ambiguity (betrayal-seeking) for small likelihoods of winning and prefer nature ambiguity over betrayal ambiguity (betrayal-averse) for larger likelihoods of winning.

TABLE 3.1 Ambiguity attitudes by ambiguity treatment

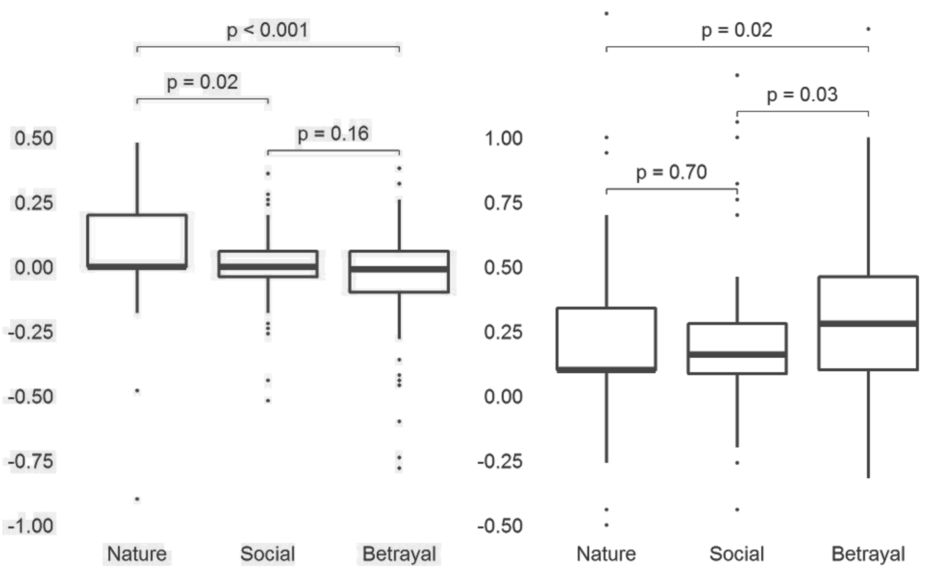
	I. Ambiguity aversion index (b)			
Ambiguity treatment	Mean	Median	Interquartile range	P-value Wilcoxon tests AA = 0
Nature	0.07	0.00	[0.00, 0.20]	0.00
Social	0.01	0.00	[-0.04, 0.06]	0.27
Betrayal	-0.05	-0.01	[-0.10, 0.06]	0.29
	II. A-insensitivity index (a)			
Ambiguity treatment	Mean	Median	Interquartile range	P-value Wilcoxon tests AI = 0
Nature	0.20	0.10	[0.10, 0.34]	0.00
Social	0.20	0.16	[0.08, 0.28]	0.00
Betrayal	0.28	0.28	[0.10, 0.46]	0.00

3.6 Discussion

We find that people's attitudes to ambiguity in strategic interactions with others are different from their attitudes toward nature-generated ambiguity that is non-social and non-strategic. Ambiguity aversion, which has been found in many preceding empirical studies and which we also find for nature-generated ambiguity, plays no role in strategic interactions. Our subjects are not averse to betrayal ambiguity. Nor did we find ambiguity aversion when subjects face social ambiguity without strategic complications. People are less averse toward ambiguity generated by other humans—regardless of strategic aspects—than non-social Ellsberg-like mechanisms. In particular, our findings contradict betrayal aversion.

FIGURE 3.7 Differences in ambiguity attitudes

I. Ambiguity aversion index (b)   II. A-insensitivity index (a)



NOTES: Box-and-whiskers plots of distributions of ambiguity aversion index (left panel) and a-insensitivity index (right panel) by ambiguity treatment. The box shows the interquartile range and the median (solid line) of each distribution. The whiskers show the lower and upper quartiles, excluding outliers. Outliers—values larger than (3/2) times the upper quartile or lower than (3/2) times the lower quartile—are indicated by dots. All tests of differences are Mann-Whitney U tests.

The competence hypothesis (Heath and Tversky, 1991) may explain our findings: people do not dislike ambiguity if they feel competent about the source generating the ambiguity. Heath and Tversky (1991) found that people preferred to bet on ambiguous events rather than equiprobable chance events when they considered themselves knowledgeable, but not otherwise. Fox and Weber (2002) confirmed the competence effect in games, analyzed as choices under ambiguity. Studies comparing preferences between ambiguous events also find that people prefer the familiar over the unfamiliar source. For instance, they prefer to bet on domestic over foreign stocks (French and Poterba 1991, Dimmock et al. 2016) and on weather events in their home country rather than abroad (Abdellaoui et al. 2014, Chew, Ebstein, and Zhong 2012, de Lara Resende and Wu 2010). In our experiment, whereas it is natural that subjects felt ignorant about the Ellsberg-like mechanism, they may have felt more competent or knowledgeable in their judgment about a fellow subject's trustworthiness or her snack preferences. Apparently, people feel a special aversion toward artificially created Ellsberg-like ambiguities that is not representative for ambiguity in general.

Whereas our subjects do not display ambiguity aversion toward social ambiguity, they still violate ambiguity neutrality. This violation is cognitive rather than motivational. There is significant likelihood a-insensitivity for all—social and non-social—ambiguities, but a-insensitivity is stronger for betrayal ambiguity. Whereas classical theories as used in B1 and B2's analysis have focused on motivational components, our findings add to the growing empirical evidence that insensitivity is an important component. A-insensitivity also has implications for the interaction between betrayal and ambiguity aversion. It implies that people prefer betrayal ambiguity over nature ambiguity (betrayal-seeking) for small likelihoods of winning but reverse these preferences for larger likelihoods of winning, in agreement with Fetchenhauer and Dunning's (2012) findings.

Our separation between social and nature ambiguity should be distinguished from the separation between ambiguity experiments with and without suspicion. In the latter, experimenters have the possibility to, for instance, rig urns to reduce subjects' earnings. Suspicion does not play any role in our experiment. In the nature experiment, for instance, subjects gamble on or against all events. The experimenters have no interest in making some events more or less likely than others. The Prince incentive system further rules out any suspicion. It can be argued that our nature treatment is also partly social, because human beings (experimenters) had designed the experiment with marked cards. This process was random though, and the experimenters did not have any deliberate interest in marking cards one way or the other. In the social treatment, human beings had interests and decided willfully. We acknowledge that our nature treatment did not completely rule out human intervention, but only reduced it relative to the social treatment. It is difficult to completely rule out any human influence in whichever situation we face in life, and surely in tasks organized by experimenters. For example, Oechssler and Roomets (2015), whose paper contributed to the development of our ideas[16], generated ambiguity by having subjects construct a Galton box where balls fell through, so as to minimize prior knowledge of probabilities by the experimenters. Here there still was human intervention, with the

[16] As did general discussions of social risk by B2. Kugler, Connolly, and Ordóñez (2012) also compared nature uncertainty with strategic uncertainty, finding differential effects of fear and anger.

experimenters planning the design and providing the facilities, and other subjects constructing the Galton boxes. The purpose of their study was to compare the Galton-box treatment with a regular treatment that did not rule out suspicion. They share with us an interest in the degree of human intervention.

3.7 Conclusion

Using a recently introduced technique for measuring ambiguity attitudes for natural events and strategic games, this paper has introduced the distinction between social and nature ambiguity. Social ambiguity, concerning willful acts chosen by interested human beings, is perceived differently than the uninterested random ambiguity coming from nature. We apply our new controls for ambiguity attitudes to shed new light on betrayal aversion. We find no betrayal aversion. Instead, our findings are as follows. People are less averse to strategic ambiguity than to nature ambiguity, but this is due to the general social rather than the specific strategic (betrayal) component. People are more insensitive to strategic than to nature ambiguity, and this is due to the strategic rather than the social component. Put differently, the motivational difference between strategic and nature ambiguity is due to the social component: people rather like ambiguity coming from other human beings than from nature, suggesting that there is more cooperation than competition between human beings. The cognitive difference between strategic and nature ambiguity is due to the strategic component: strategic uncertainty is harder to understand than nonstrategic uncertainty.

3.A Appendix

This appendix re-analyzes the MAPs, used by B1, B2, using modern ambiguity theories. We show that MAPs are more complex than classical theories would suggest. MAPs involve not only attitudes toward risk, but also attitudes toward ambiguity, complexity, and dynamic optimization under non-expected utility. It is not easy to draw conclusions from MAP measurements because many factors play a role.

FIGURE 3.A1 *MAP*s

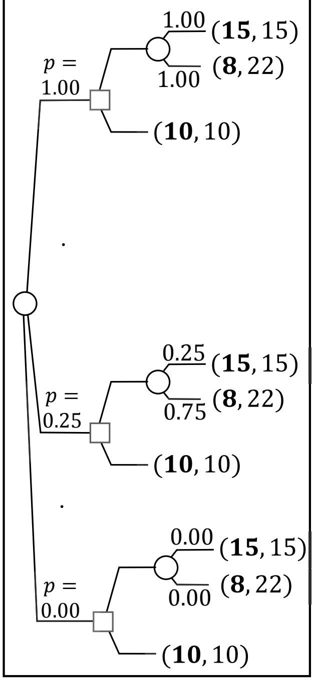
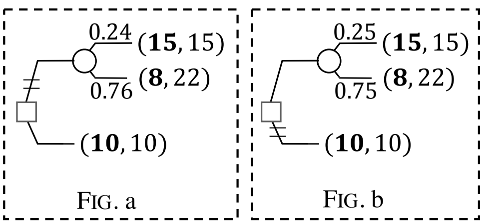


FIGURE 3A.2 Risk preferences



Consider one subject, a trustor, whose outcomes are bold—we do not denote currency; see Figure 3.A1. The subject has to choose between going up or down in the square nodes in the second stage, with one further restriction specified later. Going up gives the favorable outcome with objective probability p and the unfavorable outcome otherwise, and going down gives the intermediate outcome for sure. The probability p is determined in the first stage, as follows. In B1 and B2's Trust game treatment, the subject is informed that p refers to the percentage of subjects choosing to be trustworthy (choosing (**15**,15) in the circle node) in an experiment. The subject will be randomly matched with a subject from that experiment and then play the trust game, so that p indeed is the probability of the favorable outcome of up at the final circle node. In B1 and B2's Risky Dictator treatment[17], no information is given about how p was determined. In both treatments, the uncertainty about p, in the first circle node in Figure 3A.1, is ambiguous to the subject.

The further restriction announced above is that the subject, if choosing up for some probability p, must also do so for all higher probabilities p. That is, the subject has to choose a minimally acceptable probability (MAP) and then choose up for all $p \geq MAP$ and down for all $p < MAP$. It is likely that subjects will satisfy this restriction anyhow, so that it is not a serious restriction.

Assume a subject with the risk preferences in Figure 3A.2, where bars indicate the dispreferred options. The minimal probability $q = j/100$ with $(10,10) \preccurlyeq (q: (15,15), 1 - q: (8,22))$ is $q = 0.25$. B1 and B2 argue that a choice $MAP = 0.25$ in Figure 3A.1 then is rational, invoking a substitution axiom of expected utility. This axiom implies a separable treatment in, for instance, the branch $p = 0.25$, where the preference should be as in Figure 3A.2b, irrespective of what happens at other branches $p \neq 0.25$. This separability assumption has been discussed in some places in the literature. It underlies the random incentive system for implementing real incentives. In the Anscombe-Aumann (1963) framework of ambiguity, it received the misleading name monotonicity. Skiadas (2013 p. 63) pointed out: "This is not an innocuous assumption." Many theoretical studies criticized the condition on both normative and descriptive grounds.[18] Empirical tests falsifying it include Baltussen et al. (2012), Cox, Sadiraj, and Schmidt 2015), and Schneider and Schonger (2015). Hence, we conclude that, in the present state of the art, the assumption cannot be maintained for MAP measurements. We next turn to an analysis using modern ambiguity theories.

In the experiments, subjects perceive three different events:

(1) getting the good outcome (**15**,15), i.e. $p \geq MAP$ and up in the circle nodes;

(2) getting the middle outcome (10,10), i.e. $p < MAP$;

(3) getting the bad outcome (**8**,22), i.e. $p \geq MAP$ and down in the circle nodes.

[17] Treatments in B1 and B2 are between-subjects.

[18] See Bade (2015), Bommier (2014), Cheridito et al. (2015), Ghirardato (2002), Hammond (1988), Holt (1986), Karni and Safra (1987), Machina (1989; 2014 p. 385 3rd bulleted point), Wakker 2010 (Figure 10.7.1).

A decomposition as in Figure 3.A1 is hard to come by for subjects. Instead, the three events are perceived as complex and ambiguous. We now give a numerical example showing how ambiguity aversion, rather than betrayal aversion, can lead to the findings of B1 and B2.

Example. Assume that the subject maximizes expected utility under risk, with $U(\mathbf{8}, 22) = 0, U(\mathbf{15}, 15) = 1$, and $U(\mathbf{10}, 10) = 0.245$. This gives the risky preferences of Figure 3A.2. For ambiguity, the subject's preferences can be modeled both using Gilboa and Schmeidler's (1989) maxmin expected utility and using Schmeidler's (1989) Choquet expected utility. We use the formulas of the latter theory, abbreviated CEU, because they are more tractable.[19] The subject's weighting function W on events is of the form $w_A(P(.))$. Here P may be interpreted as subjective belief. For simplicity we assume that it is derived from given prior and posterior probabilities according to the rules of probability calculus with details given later. Because there is uncertainty about P and the subject is ambiguity averse, she scales these probabilities down by $w_A(p) = p^2$. This w_A, the *source function,* captures ambiguity attitudes. A MAP is evaluated by

$$CEU = w_A\big(P(\mathbf{15}, 15)\big) \times 1 + \Big(w_A\big(P(\mathbf{15}, 15) + P(\mathbf{10}, 10)\big) - w_A\big(P(\mathbf{15}, 15)\big)\Big) \times 0.245 \quad \text{(A.1)}$$

where $P(\mathbf{15}, 15)$ is the probability of receiving the best outcome with utility 1 and $P(\mathbf{10}, 10)$ is the probability of receiving the middle outcome with utility 0.245. We suppress the term of utility 0 that the worst outcome contributes.

In the risky dictator game, the subject considers all p equally probable, assigning prior probability $\pi(p) = 1/101$ to each. With $MAP = 0.25, P(\mathbf{10}, 10) = 25/101$ and $P(\mathbf{15}, 15) = \sum_{j=25}^{100} \left(\frac{1}{101}\right) \times \frac{j}{100} = 0.4703$. The CEU value of this MAP is 0.293. The best MAP is considerably larger, i.e., 0.39, with CEU value 0.299269. That the best MAP deviates from 0.25 shows that ambiguity aversion can imply violations of B1 and B2's substitution/separability assumption.

Assume that in the Trust game treatment the subject received a signal of $p = 0.29$.[20] She believes that the signal is correct with probability 0.5. In the other case, she assigns

---

[19] Because the function w_A defined below is convex, so is the event weighting function $w_A(P)$ defined below (Wakker 2010 Exercise 10.9.1a and p. 448). Hence the CEU model that we use (Eq. A.1) is also a maxmin EU model (Wakker 2010 Theorem 11.4.1).

[20] B2 found $p = 0.29$ for their two largest countries: China and the US.

uniform prior probabilities $1/101$ to all p. Hence, $\pi(p = 0.29) = 0.5 + 0.5/101$ and the remaining probability is divided uniformly over all other p. $CEU(MAP = 0.25) = 0.1713$. Again, the best MAP is considerably larger, i.e. 0.73, with CEU value 0.246771. It also considerably exceeds the best $MAP = 0.39$ found in the Risky dictator treatment. B1 and B2 interpreted the latter discrepancy between the two treatments as betrayal aversion, but in this example it is merely due to ambiguity aversion. Note that in this hypothetical example not ambiguity aversion per se, but a different effect of it in the two treatments drove our results.

In the example the higher optimal MAP for the Trust game treatment than for the Risky dictator treatment was caused by different subjective beliefs, which made the same ambiguity attitude (w_A) have more effect in the former than in the latter treatment. We chose models with ambiguity aversion because they have been best known. Many other examples can be constructed with different ambiguity attitudes in the two treatments or different non-expected-utility evaluations of the dynamic optimization problem. There are many components besides betrayal aversion and no easy conclusions can be drawn from the complex MAP measurements.

Quercia (2015) replicated the experiment of B1 and B2 but made three simplifications to the MAP design. First, she derived MAPs indirectly from choice lists, second, she expressed probabilities in terms of frequencies, and, third, she improved the instructions. These improvements led to better understanding but not to different results. They do not resolve the problems discussed in this appendix.

Chapter 4

Communication and Cooperation in a High Stakes TV Game Show

with MARTIJN J. VAN DEN ASSEM AND DENNIE VAN DOLDER

4.1 Introduction

In many social interactions people care about the intended behavior of others. Often, the information that is available for predicting what others will do is limited to non-binding and non-verifiable communication. Under the neo-classical assumption that lying is costless, such cheap talk is not informative when incentives are insufficiently aligned (Farrel and Rabin 1996). This assumption, however, contradicts the view held by many in social psychology that for most people lying entails an unpleasant—therefore, costly—experience (Eckman 2001; Vrij 2008). An extensive body of research on cues to deception critically builds on the assumption that lying and truth-telling generate distinct emotional experiences (Zuckerman et al. 1981; DePaulo et al. 2003; Sporer and Schwandt 2007). In line with psychologists' view, numerous studies have shown that pre-play communication—especially "promises"—enhances cooperative behavior (Sally 1995; Balliet 2010). The notion that people are unscrupulous liars is also rejected by a recent and growing literature in experimental economics (Gneezy 2005; Lundquist et al. 2009; Serra-Garcia et al. 2011; Erat and Gneezy 2012; Cappelen et al. 2013).

There are a number of explanations for people's aversion to lying. Commitment-based explanations posit that people simply have an intrinsic preference for keeping their word (Ellingsen and Johannesson 2004; Vanberg 2008). Such a preference, possibly shaped by social norms (Krupka et al. 2016), is in line with self-concept maintenance theory, where people like to view themselves as honest and are aversive to negatively updating their self-concept after a dishonest act (Mazar et al. 2008). Alternatively, expectation-based accounts state that people dislike lying because they experience guilt if they do not live up to others' expectations (Charness and Dufwenberg 2006, 2010; Battigalli et al. 2013). Finally, people may simply refrain from lying because of fear of being caught, which could harm their reputation as an honest and reliable person.

The literature on lying aversion and the predictive power of promises has concentrated on binary—to lie or not to lie—stylized choice contexts. When talk is free-form, as in most real-life situations, the set of possible deceptive statements is richer: instead of outright lying, people can also choose to deceive by omitting, obfuscating or stretching the truth. This wider choice set is potentially important as some types of deception may be more aversive to the liar than other types, and rather be avoided. Consequently, some statements may be more indicative of cooperative behavior than other statements. Although several papers study the predictive power of statements made in free-form communication, they typically distinguish only one particular type of message (a promise) so that cheap talk is effectively still analyzed in a binary choice framework (Charness and Dufwenberg 2006; Vanberg 2008; Belot et al. 2010; van den Assem et al. 2012).

In the present paper, we empirically address the question whether the distinction between different types of statements adds to the predictive power of cheap talk using a high-stakes non-cooperative game with free-form communication. We argue that lies are less costly if they are malleable to interpretation as truths. We hypothesize that people who defect prefer to make statements that allow them to deny the fact that they are lying. Such statements arguably entail weaker commitment, weaker feelings of disappointing the other, and lower reputation costs.

We propose a typology of statements in terms of their malleability to interpretation as truths. Using data from the TV game show Golden Balls we examine to what extent the different types are informative of subsequent decisions.[21] In each episode of this show, two contestants play a variant on the Prisoner's Dilemma where they simultaneously decide to either split (cooperate) or steal (defect) a sum of money that on average exceeds $20.000. Prior to their decisions, they briefly engage in a free-form discussion about the choice at hand. During the talk, they typically exchange multiple statements, most of which involve giving or eliciting some type of signal that the intended decision is to split.

[21] Game show data have been widely used in economics to study, for example, individual decision-making under risk (Gertner 1993; Metrick 1995; Beetsma and Schotman 2001; Post et al. 2008; Hartley et al. 2013), strategic decision-making (Bennett and Hickman 1993; Berk et al. 1996; Tenorio and Cason 2002), discrimination (Levitt 2004; Antonovics et al. 2005), cooperative behavior (List 2004, 2006; Belot et al. 2010; Oberholzer-Gee et al. 2010; van den Assem et al. 2012) and bargaining (van Dolder et al. 2015).

Our typology distinguishes between a contestant's signals according to two dimensions. First, it discriminates between statements explicitly expressing that the contestant will choose to split and statements that only implicitly signal that the contestant will do so. Second, it discriminates between unconditional statements and statements that carry an element of conditionality on what the other will do. We argue that explicit and unconditional statements are less malleable to denial than implicit or conditional statements. Consider, for example, the statement "I will split". This statement is both explicit and unconditional, and a defector who used it will find it hard to deny that she has deceived her opponent. The statement "I came here to split" also has no element of conditionality, but this one is at best only an implicit promise to split: it is silent about the contestant's current intention and she may have changed her mind. The explicit statement "I will split if you split" is clearly conditional on the opponent's choice, and a decision to steal can be justified by a pretended belief that the opponent also steals.

Earlier studies providing support for our malleability hypothesis include Schweitzer and Hsee (2002) and Serra-Garcia et al. (2011). Schweitzer and Hsee (2002) find that subjects were more likely to overstate or understate the value of a good in a selling or buying context when their private information was more uncertain or vague. Serra-Garcia et al. (2011) find that subjects prefer to deceive by means of a vague message that also captures the truth over deceiving by outright lying.

Consistent with the malleability hypothesis, we find that unconditional statements predict a higher likelihood of cooperation than conditional statements, and that making an unconditional and explicit statement predicts a higher likelihood of cooperation than making an unconditional statement that is implicit. In our data, making an unconditional and explicit statement ("I will split") is associated with a 22 %age point increase in the likelihood that the contestant will in fact choose the split ball. Making an unconditional and implicit statement ("I want to split", "I came to split") is associated with an increase of 12 %age points in the contestant's likelihood of choosing split. Conditional statements have no predictive power. Furthermore, we find that malleability of the contestant's statements, rather than their quantity, predicts her decision. These findings are robust after controlling for observable contestant and game characteristics, such as the contestant's gender and age, level of education, the size of the stake at play, and a proxy for the contestant's familiarity with the show.

Our study relates to a broader literature on deception. Lundquist et al. (2009) and Hilbig and Hessler (2012) show that people are more willing to tell small rather than big lies. Gneezy (2005) and Erat and Gneezy (2012) show that people care about the consequences of a lie for both themselves and for others. The more a person can gain from lying, the more likely she is to lie. Alternatively, the more a lie hurts the other person, the less likely a person is to lie. Studying behavior in a TV game show similar to Golden Balls, Belot et al. (2010) find that people are more likely to stick to a promise if it was made voluntarily than if it was elicited by a question from the host. Similarly, Charness and Dufwenberg (2006, 2010) show that people have a greater tendency to stick to promises in free-form communication, compared to a situation where they can only send either an explicit promise or no message at all. Shalvi et al. (2011) show that normatively irrelevant aspects of the decision context can provide subjects with a justification for lying. Finally, research on moral judgment has found that people view deception by omission as less immoral than deception through commission (Spranca et al. 1991). This suggests that people will also be more willing to deceive through omission than through commission themselves.

4.2 Description of game show and data

4.2.1 Golden Balls

Golden Balls was broadcast on British television between June 2007 and December 2009. We analyze the game played in the fourth and final round of each episode. In this final, two contestants play a game that resembles the classical Prisoner's Dilemma. What follows here is a brief description of the game. For a more extensive discussion we refer to van den Assem et al. (2012).

The show begins with four contestants who have not met before. In the first two rounds, the four (round 1) or remaining three contestants (round 2) all receive a set of golden balls. Each ball carries a specific value, which in the end may contribute to the final jackpot. The contents of some balls are visible to everyone, while the contents of the other balls are known to the contestant only. The contestants' task is to make claims about the contents of their hidden balls, after which they have an open discussion and then cast votes against each other. The player who receives most votes is eliminated from the game, together with her golden balls. In the third round, the two remaining contestants determine the final jackpot through a random draw from the remaining balls.

In the fourth and final round, each of the two finalists is presented with two golden balls, one with the word "split" and the other with the word "steal" written inside. They simultaneously have to choose either the split or the steal ball. If both decide to split, they split the jackpot equally. If one decides to split while the other decides to steal, the one who steals receives the entire jackpot and the one who splits goes home with nothing. If both decide to steal, both go home with nothing. Prior to their decisions, the two contestants briefly engage in a free-form talk about the choice at hand.

4.2.2 Van den Assem et al. (2012)

Golden Balls has previously been studied by van den Assem et al. (2012). In the analyses in the next section we control for the determinants of cooperation identified in that study.

Van den Assem et al. (2012) report that young males are less cooperative than young females, and that this gender effect reverses for older contestants because men become more cooperative as their age increases. In addition, they find some evidence that white and higher educated contestants are more likely to cooperate, and that distinguishing between contestants from urban and rural areas and between students and non-students adds little explanatory power.

Cooperation is largely insensitive to the size of the jackpot. A major exception concerns the smallest stakes, for which the rate of cooperation is relatively high. Cooperation rates are also influenced by the maximum potential jackpot, a value that is being emphasized prior to the final. The higher this maximum is and the smaller the actual jackpot thus appears, the greater the likelihood that players cooperate. This effect diminishes with the number of televised episodes. Altogether, these results suggest that relative thinking plays a role here.

Consistent with the notion that people have a preference for reciprocity, contestants are less likely to cooperate with opponents who have tried to vote them off the show during the first two rounds. Contestants who earlier on misrepresented the contents of their hidden balls do not cooperate significantly more or less than those who have been honest throughout, and they do not face more or less cooperative choices by their opponents in the final.

The prior study also includes a crude analysis of the predictive power of cheap talk, and reports that those who make an unambiguous promise are much more likely to cooperate. Finally, van den Assem et al. (2012) find little evidence that contestants' propensity to cooperate depends on the likelihood that their opponent cooperates: players do not seem to condition on their opponent's promise or background characteristics.

4.2.3 Typology of pre-play communication

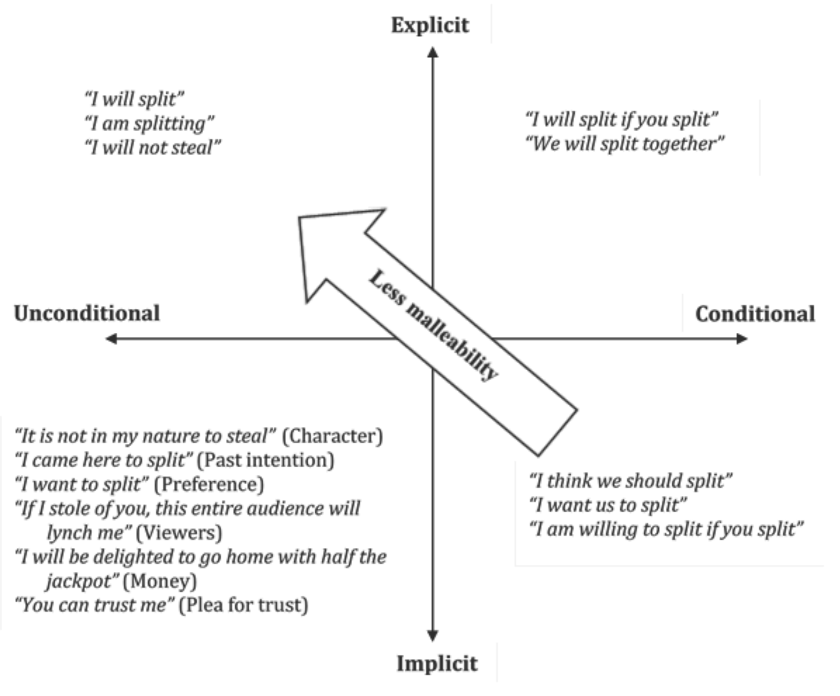
Prior to their split or steal decisions, the two finalists briefly engage in a free-form talk about the choice at hand. During the talk, the contestants typically exchange multiple statements through which they try to signal that they will choose to split or to get assurances that the opponent will do so. Owing to the free-form nature of the talk, there is a rich set of statements that contestants can make. In this subsection we introduce the typology of cheap talk that is central in the present paper.

We categorize the different signal types according to how malleable these are. Contestants who are about to steal want to conceal this. At the same time, they are shy of lying. Based on the notion that lies that are more malleable to interpretation as truths entail lower lying costs, we expect that stealing contestants will prefer malleable statements over unmalleable statements. Contestants who are about to split, however, face no lying costs and thus have no need for such substitution. As a consequence, the malleability of the statements that contestants make predicts cooperative behavior.

As explained in the introduction, our typology distinguishes between explicit and implicit statements and between unconditional and conditional statements. These distinctions lead to four different signal types. Figure 4.1 gives a schematic overview and typical examples of statements of each type.

Explicit and unconditional statements such as "I will split" are completely unmalleable. Consider, in contrast, the statement "I want to split". The latter does not explicitly express that the contestant will choose split, but signals it by expressing a preference to do so. Strictly speaking, the desire to split can be sincere even if the action is to the contrary, and therefore we argue that this statement is more malleable than the former. In a similar way, a contestant can signal that she will split without stating it explicitly by referring to her character ("It is not in my nature to steal"), referring to a past intention ("I came to split"), expressing awareness or worry that viewers will disapprove stealing ("If I stole off you, this entire audience will lynch me"), indicating that half the jackpot is good enough or a lot of money ("I will be delighted to go home with half the jackpot"), and by urging the opponent to have trust ("You can trust me"). All these subtypes of unconditional and implicit statements allow a contestant to deny that she has lied if she steals.

FIGURE 4.1 Typology of signals



Next, compare the explicit and unconditional statement "I will split" with the statement "I will split if you split". Even though both statements are explicit about what the contestant will do, the latter carries a clear element of conditionality on the opponent's choice through an if-clause. Similarly, a statement such as "We will split together" explicitly expresses what the contestant will do, and here the conditionality is implied by a reference to the coordinated action. The conditionality present in explicit and conditional statements makes it easy for a stealing contestant to deny that she has lied, as she can always argue that she did not believe that her opponent was going to split. The fourth type comprises statements that are both implicit and conditional. These two properties make it easy to deny lying if one decides to steal. Examples include "I think we should split", "I want us to split," and "I am willing to choose split if you split".

Our typology only considers statements that patently signal that the contestant will choose split. Other statements are considered as empty talk. Empty talk includes questions and pleas directed at the opponent ("Will you split?", "I hope you will split"), remarks about the previous rounds ("I did not lie a single time", "I brought you to the final"), and many other types of idle remarks ("This is a lot of money", "You seem like a nice person", "We have come so far").

4.2.4 Coding rules

Coding free-form communication into the types described above is not straightforward. People stumble over words, jump from one topic to the next, and use short and incomplete phrases that are difficult to interpret. We therefore use a strict set of rules.

First, we require that statements have a meaning on their own. This implies that we ignore short utterances such as "Yeah", "No", and "Absolutely". Short utterances are typically either preceded or followed by complete statements that elaborate their meaning. These complete statements are coded, so usually no information is lost. Many short utterances constitute responses to remarks or questions by the opponent. Although such utterances have no stand-alone interpretation, their formal meaning is normally easy to infer. We nevertheless consider such responses as empty. The reason is that in a conversation between two people, people are more or less expected to respond in a particular way to what the other has said or asked. If a contestant chooses not to back up her short response with a longer statement, this can be seen as an indication that the short reply was forced out of her and that it is to be seen as empty.

Second, we count consecutive statements of the same type as one statement from that (sub)type if they are not interspersed with other types of statements. Contestants often repeat the same statement or elaborate on the same point in another way ("It is not in my nature to steal. I just could not do that."). When multiple statements of the same type are interspersed with other types, we do treat them as multiple statements as this indicates that the contestant consciously decided to go back to making this point.

Third, if the stand-alone interpretation of the verb "to split" is ambiguous, we assume that it refers to the contestant's action of choosing the split ball. Formally, the verb can refer both to the action of a single contestant who chooses the split ball and to the outcome of two contestants who share the jackpot. Often the exact meaning is clear, but for some statements, such as "I want to split", one can argue both ways. Because the more common meaning in the show is the unilateral act of picking the split ball (the final game is even called "Split or Steal?"), we use this meaning as the default interpretation.

Fourth, just like "you and I" is equal to "we", we interpret the combination of a statement referring to "you" (the opponent) and a similar statement referring to "I" (the player herself) as one statement referring to "we". For example, "You will split, I will split" is considered equivalent to "we will split". The order of the two statements does not matter, but we do require that they directly follow each other.

Last, if a statement contains a pronoun such as "it", "this" or "that" and if the proper noun to which it refers is clear from the context, then we interpret such a statement as if the proper noun replaces the pronoun.

4.2.5 Data

The data set used for this study covers 282 episodes. Recordings were originally provided by Endemol's local production company Endemol UK for the purpose of the study by van den Assem et al. (2012). Of the 288 episodes that aired in total, one could not be located by the producer and is therefore missing. We intentionally leave out five other episodes. In one of these, a finalist promises his opponent that he will pick his steal ball, and that he will reward him with half of the money after the show if the opponent splits. Though interesting in its own respect, this exceptional case is incompatible with our coding framework. In two episodes, most of what contestants say is clearly meant ironically, apparently owing to the small jackpot sizes (£3.00 and £3.65), and not to be taken seriously. Finally, in two episodes conversations were in most parts inaudible.

To construct our communication data, we first transcribe contestants' conversations, then count for each contestant the number of times she uses each of the four signal types, as well as each of the subtypes of the unconditional and implicit signal. Table 4.1 (Panel A) shows our count data. In general, it can be seen from the frequencies of signal counts that repetitions of signals were relatively rare, with hardly any contestants repeating the same signal type more than twice. In our main analysis we therefore employ binary variables that indicate for each signal type whether a contestant uses it at least once during her pre-play talk with the opponent.[22]

The binary indicators used in our main analysis are shown in the right panel of Table 4.1 (Panel B). The first two columns show the frequencies of the indicator values, and the

[22] In subsequent analyses, we also investigate the extent to which repetitions of signals predict contestants' likelihood of splitting.

third column reports the means of the binary indicators. As shown in the table, the unconditional and explicit signal statement ("I will split") was made by over one-third of all contestants. Unconditional and implicit signals were made by more than half of the contestants. Table 4.1 also shows the summary statistics for the subtypes of the unconditional and implicit signal. When making an unconditional and implicit statement, most contestants indicated that the money amount after splitting was good enough for them (27 % of all contestants), that they had a preference for choosing the "split" ball (20 % of all contestants), or that they had planned all along to choose the "split" ball (14 % of all contestants). Conditional statements were made less often, with only 7 % of the contestants having made a conditional and explicit statement ("I will split if you split"), and 32 % of the contestants having made a conditional and implicit statement ("I think we should split", "I want us to split").

TABLE 4.1 Frequencies of signal counts and binary indicators

Signal type	Panel A Frequencies of signal counts							Panel B Signal indicators		
	0	1	2	3	4	5	6	0	1	Mean
Unconditional and explicit	353	145	50	12	1	1	2	353	211	0.37
Unconditional and implicit	248	185	90	30	6	3	2	248	316	0.56
Character	537	23	3	1	0	0	0	537	27	0.05
Past intention	484	67	12	1	0	0	0	484	80	0.14
Preference	452	95	14	3	0	0	0	452	112	0.20
Viewers	529	30	5	0	0	0	0	529	35	0.06
Money	411	132	21	0	0	0	0	411	153	0.27
Plea for trust	532	30	2	0	0	0	0	532	32	0.06
Conditional and explicit	525	34	4	1	0	0	0	525	39	0.07
Conditional and implicit	384	140	34	6	0	0	0	384	180	0.32

4.3 Results

4.3.1 Descriptive results

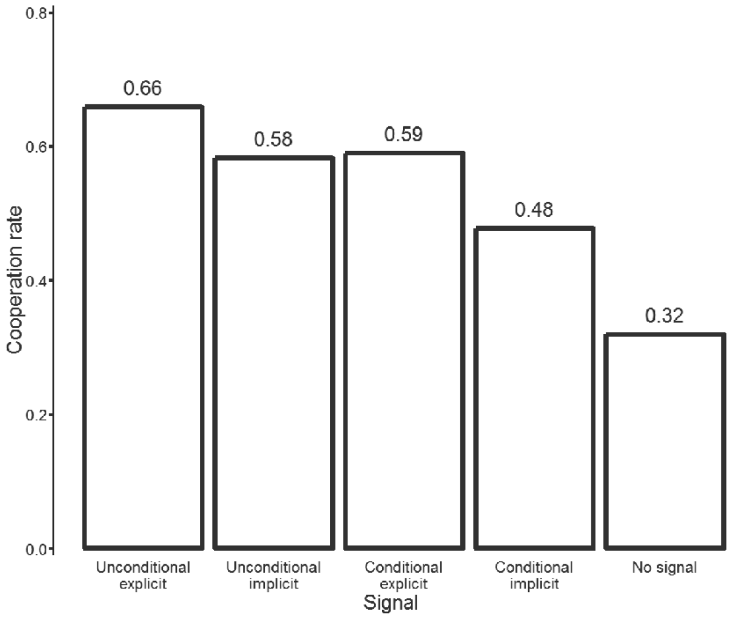
Our main hypothesis is that malleability of contestants' signals predicts their likelihood of splitting. Specifically, we expect the least malleable unconditional and explicit signal to have the highest cost of lying and therefore the strongest influence on the contestant's splitting likelihood. We expect the most malleable conditional and implicit signal to have the lowest cost of lying and the weakest influence on the likelihood of splitting. Finally, we expect the two signal types—the unconditional and implicit and the conditional and explicit—in the middle of our malleability-based ranking to also fall in the middle in terms of their effects on the contestant's likelihood of splitting.

Figure 4.2 shows the proportions of contestants who chose to split, depending on the types of signals that contestants gave to their opponents during the pre-play communication. The figure shows that, in line with our malleability hypotheses, of the contestants who made the unconditional and explicit signal statement 66% subsequently chose to split the jackpot. Cooperation rates were lower among contestants who gave the more malleable unconditional and implicit (58%) and conditional and explicit (59%) signals. Of the contestants who gave the most malleable conditional and implicit signal, 48% chose to split. Finally, contestants who gave no signals (made empty statements only) had the smallest likelihood of splitting (32%). Thus, the pattern shown in Figure 4.2 is consistent with our hypothesis that the malleability of cheap talk statements determines their power to predict people's subsequent decisions. In the following sections, we analyze the relationship between contestants' signals and their decisions to split or steal using the logit choice model.

4.3.2 Estimation strategy

To estimate how a contestant's decision to split is explained by her signals we assume that following the pre-play talk she chooses to split if the (expected) utility of doing so exceeds the utility of stealing. Furthermore, we assume that her utility of choosing split is a function of her demographic characteristics X and the game's context characteristics Z, $U^{split} = \alpha^{split}X + \beta^{split}Z + \varepsilon^{split}$. Similarly, the utility of choosing steal is a function of the demographic and context characteristics. In addition, we assume that the utility of stealing is diminished by the cost of lying, which the contestant experiences if she had signaled to split but decides instead to steal, $U^{steal} = \alpha^{steal}X + \beta^{steal}Z - \gamma S + \varepsilon^{steal}$. The stronger the signals S that she gives to her opponent the higher is the cost experienced for deciding to steal, and the less attractive stealing becomes. Finally, ε^{split} and ε^{steal} capture the components of the contestant's utilities of splitting and stealing that are not accounted for by the deterministic functions.

FIGURE 4.2 Cooperation rates by signal type



Assuming that the unobserved components are *i.i.d.* extreme value leads to the estimation of the contestant's likelihood of splitting as $P^{split} = 1/[1 + \exp(\alpha X + \beta Z - \gamma S)]$, where the parameters $\alpha = \alpha^{steal} - \alpha^{split}$ and $\beta = \beta^{steal} - \beta^{split}$ capture the utility advantage to the contestant of stealing compared to splitting. In our analysis we specify $\alpha X + \beta Z$ in the same way as the basic model estimated by Van den Assem et al. (2012, Model 6), which includes all the aforementioned (Section 4.2.2) demographic and context characteristics found to be related to the split or steal decision of the contestants. Throughout our analyses we keep the specification of $\alpha X + \beta Z$ fixed, and refer to its variables as controls. The focus of this paper is the set of cost of lying parameters γ associated with the contestants' signals S.

Parameters in our analyses are estimated with the maximum likelihood method. In estimating the standard errors, we account for possible correlations between the decisions of contestants within the same episode and apply a clustering correction (see, for example, Wooldridge 2003). To facilitate interpretation of the estimates, we report the signal variables' (average) marginal effects. Standard errors of the marginal effects are calculated using the (first-order) delta method (see, for example, Casella and Berger 2002).

4.3.3 Estimation results

Malleability of signals and their credibility

Our basic model (Model 1) specifies the cost of lying as the weighted sum of the lying costs associated with each signal type. Here, the coefficients of the indicators of signal types capture the magnitudes of lying costs associated with each signal type.

Table 4.2 (Models 1 and 2) shows for each signal type the marginal effect of making this type of a signal statement (at least once) on the contestant's likelihood of splitting. Model 1 does not include controls for demographic and context characteristics, Model 2 includes the full set of controls. The marginal effects of the contestant's signals remain essentially unchanged after controlling for her demographic and context characteristics.

We find that if during the pre-play talk the contestant gives the unconditional and explicit signal that she will split (e.g., "I will split"), she is 22 percentage points more likely to actually choose split compared to a contestant who does not make such a statement. The unconditional and implicit signal (e.g., "I want to split", "I came to split") also predicts a higher likelihood of splitting. If the contestant makes such a signal statement, she is 12 percentage points more likely to choose split. By contrast, conditional signals (e.g., "I will split if you split", "I want us to split") contain no information about the contestant's subsequent decision, regardless of whether these are explicit or implicit.

A test of equal effect sizes rejects the null hypothesis that these four effects are equal in magnitude (Wald test, p-value < 0.001). Pairwise tests show that the unconditional and explicit signal was more predictive of behavior than the unconditional and implicit signal (one-sided p-value = 0.045), the conditional and explicit signal (p-value = 0.031), and the conditional and implicit signal (p-value < 0.001). The effects of the two signals in the middle of our malleability-based ranking—the unconditional and implicit signal and the conditional and explicit signal—do not differ significantly from each other (two-sided p-value = 0.470). When compared to the conditional and implicit signal, the unconditional and implicit signal was more predictive of behavior (one-sided p-value = 0.004), and the conditional and explicit signal did not differ significantly (one-sided p-value = 0.169).

TABLE 4.2 Effects of signal types on probability of choosing “split”

	MODEL 1	MODEL 2	MODEL 3
Dependent variable	Decision (split = 1)	Decision (split = 1)	Decision (split = 1)
Estimated model	Logit	Logit	Logit
Signal type	Average marginal effect (Standard error)	Average marginal effect (Standard error)	Average marginal effect (Standard error)
Unconditional and explicit	0.199***	0.224***	0.212***
	(0.044)	(0.042)	(0.042)
Unconditional and implicit	0.109**	0.121**	
	(0.043)	(0.041)	
Character			0.238***
			(0.077)
Past intention			0.161***
			(0.055)
Viewers			0.165**
			(0.085)
Preference			0.125***
			(0.046)
Money			-0.016
			(0.042)
Plea for trust			-0.017
			(0.087)
Conditional and explicit	0.069	0.054	0.060
	(0.079)	(0.080)	(0.078)
Conditional and implicit	-0.039	-0.031	-0.022
	(0.046)	(0.044)	(0.043)
Controls	No	Yes	Yes
Observations	564	564	564
Clusters	282	282	282
Wald chi2	29.51	94.37	100.48
(df, p-value)	(4, 0.000)	(17, 0.000)	(22, 0.000)

*** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$.

In Model 3 we break down the unconditional and implicit signal into its multiple subtypes, and replace its indicator with six separate indicators of whether the contestant makes each of the subtypes of the unconditional and implicit signal. In Model 1 the indicator of the unconditional and implicit signal takes the same value regardless, for example, of whether the contestant gives the signal by referring to her character or her preference. The estimates reveal that the predictive power of signals varies significantly within the unconditional and implicit type (Wald test, p-value = 0.015).

The estimates show that some implicit signals are highly predictive of behavior, with the effect sizes comparable to that of the unconditional and explicit signal. We find that when the contestant's implicit statement refers to her character ("It is not in my nature to steal"), she is on average 24 percentage points more likely to choose split. Implicit statements referring to the contestant's past intention ("I came here to split", "I was always going to split") predict a 16 percent higher likelihood of splitting, and when the contestant refers to her concern about viewers' disapproval of stealing ("If I stole off you, this entire audience will lynch me"), she is 17 percentage points more likely to choose split. When the contestant makes an implicit statement that refers to her preference for splitting ("I want to split"), she is 13 percentage points more likely to split. Finally, implicit statements that refer to the money amount ("I will be happy to go home with half the jackpot") and that urge the opponent to have trust ("You can trust me") turn out to have no effect on the contestant's likelihood of actually choosing split.

Pairwise tests comparing the effect size of each to that of the unconditional and explicit signal reveal that the implicit signals about character (one-sided p-value = 0.327), past intention (one-sided p-value = 0.274), and viewers' disapproval (one-sided p-value = 0.359) are not less predictive than the explicit signal. However, when the implicit signal refers to the contestant's preference, the effect is smaller (one-sided p-value = 0.091).

Models 1-3 show for each signal (sub)type the effect of making the signal statement at least once. In Model 4 we expand the analysis by accounting additionally for differences in the quantities of signals that contestants give. We account for repeated signals by adding for each signal type an indicator of whether the signal is given repeatedly (twice or more times). Table 4.3 shows the results. Here, for each signal type we calculate the marginal effect of making the signal statement the first time and the (additional) marginal effect of repeating it. As can be seen from the table, regardless of the type of the signal, repetitions have no additional effects on the contestant's likelihood of splitting.

TABLE 4.3 Effects of signals and repeated signals on probability of choosing "split"

	MODEL 4
Dependent variable	Decision (split = 1)
Estimated model	Logit
Signal type	Average marginal effect (Standard error)
Unconditional and explicit	
First signal	0.211***
	(0.045)
Additional signal	0.050
	(0.061)
Unconditional and implicit	
First signal	0.108**
	(0.046)
Additional signal	0.032
	(0.052)
Conditional and explicit	
First signal	0.066
	(0.085)
Additional signal	-0.026
	(0.203)
Conditional and implicit	
First signal	-0.056
	(0.048)
Additional signal	0.118
	(0.082)
Controls	Yes
Observations	564
Clusters	282
Wald chi2 (df, p-value)	96.01 (21, 0.000)

*** p < 0.01; ** p < 0.05; * p < 0.1.

Opponent's signals

A question that naturally follows the results of the preceding analyses concerns the extent to which the opponent's signals affect the contestant's likelihood of splitting. In Model 5 we investigate this issue by adding to our analysis indicators of the types of signal statements that the opponent makes during the pre-play talk (Table 4.4). The results show that, while the contestant is more likely to split if she herself gives unmalleable signals, she is no more likely to split if the opponent signals splitting.

TABLE 4.4 Opponent's signals

	MODEL 5
Dependent variable	Decision (split = 1)
Estimated model	Logit
Signal type	Average marginal effect (Standard error)
Unconditional and explicit	0.230***
	(0.042)
Unconditional and implicit	0.125***
	(0.040)
Conditional and explicit	0.050
	(0.082)
Conditional and implicit	-0.033
	(0.045)
Opp. Unconditional and explicit	-0.005
	(0.044)
Opp. Unconditional and implicit	-0.025
	(0.040)
Opp. Conditional and explicit	0.025
	(0.077)
Opp. Conditional and implicit	0.058
	(0.044)
Controls	Yes
Observations	564
Clusters	282
Wald chi2 (df, p-value)	95.82 (21, 0.000)

*** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$.

Models 6 and 7 explore the possibility that the opponent's signals may interact with the contestant's own signals so that the opponent's signals matter indirectly by affecting the credibility of the contestant's own signals. To this end we interact in Model 6 the opponent's unconditional and explicit signal with each of the contestant's four signal types, while controlling for the opponent's other non-interacted signals. We then estimate the marginal effects of the contestant's signals both when her opponent makes an unconditional and explicit statement and when the opponent does not make such a statement. Table 4.5 (Model 6) reports the estimates. We find that the predictive power of the contestant's unconditional and explicit signal is the same (two-sided p-value = 0.727) regardless of whether her opponent gives such a signal or not. The opponent's unconditional and explicit signal also does not interact with the credibility of the contestant's unconditional and implicit signal (two-side p-value = 0.446). However, our data show that the contestant's conditional and explicit signal is more predictive of splitting if her opponent makes an unconditional and explicit statement to her than if the opponent does not make such a statement (two-sided p-value = 0.047). The predictive power of the contestant's conditional and implicit signal is the same (two-side p-value = 0.149) regardless of the opponent's signal.

In Model 7 we examine the interactions of the contestant's signals with the opponent's unconditional and implicit signal, while again controlling for the opponent's other non-interacted signals. Here, the unconditional and explicit signal turns out to be less predictive of splitting if the opponent gives an unconditional and implicit signal than if the opponent does not give such a signal (two-sided p-value = 0.004). The predictive power of the contestant's unconditional and implicit signal does not interact significantly with the opponent's such signal (two-sided p-value = 0.408). However, the contestant's conditional and explicit signal is more predictive of splitting if the opponent gives an unconditional and implicit signal (two-sided p-value = 0.072). We find that the conditional and implicit signal is less predictive of splitting if the opponent makes an unconditional and implicit statement than if the opponent does not make such a statement. As conditional statements were relatively rare, we do not have sufficient observations to analyze how the opponent's conditional signals interacted with each of the contestant's signals.

Thus, while we find evidence that the opponent's unconditional signals interact with the credibility of the contestant's own signals, the interacting directions are mixed.

TABLE 4.5 Interactions with opponent's signals

	MODEL 6		
Dependent variable	Decision (split = 1)		
Estimated model	Logit		
Signal type	Opp. Unconditional and explicit		Wald test Chi2 (p-value)
	Yes	No	H0: Yes = No
Unconditional and explicit	0.255***	0.225***	0.12
	(0.067)	(0.052)	(0.727)
Unconditional and implicit	0.088	0.149***	0.58
	(0.066)	(0.049)	(0.446)
Conditional and explicit	0.230**	-0.055	3.95
	(0.104)	(0.103)	(0.047)
Conditional and implicit	0.053	-0.076	2.09
	(0.076)	(0.053)	(0.149)
Controls	Yes		
Observations	564		
Clusters	282		
Wald chi2 (df, p-value)	98.01 (25, 0.000)		

	MODEL 7		
Dependent variable	Decision (split = 1)		
Estimated model	Logit		
Signal type	Opp. Unconditional and implicit		Wald test Chi2 (p-value)
	Yes	No	H0: Yes = No
Unconditional and explicit	0.140***	0.359***	8.46
	(0.052)	(0.059)	(0.004)
Unconditional and implicit	0.093*	0.162**	0.69
	(0.053)	(0.063)	(0.408)
Conditional and explicit	0.169*	-0.112	3.24
	(0.104)	(0.113)	(0.072)
Conditional and implicit	-0.136**	0.076	5.89
	(0.061)	(0.061)	(0.015)
Controls	Yes		
Observations	564		
Clusters	282		
Wald chi2 (df, p-value)	101.83 (25, 0.000)		

*** p < 0.01; ** p < 0.05; * p < 0.1.

4.4 Conclusion and discussion

In this paper we use the data from Golden Balls to analyze the credibility of cheap talk. Using the free-form conversations in our data, we find that different types of statements signaling future action have varying levels of predictive strength. Specifically, we conclude that malleability of a statement to ex-post justification is a good criterion for judging the likelihood that the given "promise" will be kept. The experimental literature to date has mostly used a dichotomous approach to signals, classifying them as either "promises" or "empty statements". We hypothesize that statements which are more malleable to ex-post self-serving reinterpretation can be easier for people to diverge from than statements that are unmalleable to such ex-post reinterpretation.

In the context of Golden Balls, we argue that unconditional statements of intention may be less malleable to justification than conditional statements, and that explicit statements may be less malleable than implicit statements. Our data show that controlling for other statements of a contestant, making an unconditional and explicit statement predicts a higher likelihood of splitting. The unconditional and implicit statement also predicts a higher likelihood of splitting, but to a lesser extent. Conditional statements are unrelated to probability of splitting. The differences in the predictive strengths of the statements are consistent with our hypothesis that contestants may find it more aversive to diverge from statements that are unmalleable to ex-post justification.

We conclude with comments about the generalizability of these findings and their implications for future research. First, our finding that contestants' unmalleable statements are consistent with their subsequent action does not differentiate between the alternative explanations for why lying is costly. In particular, our conceptualization of lying costs based on malleability does not exclude reputation concerns, which are explicitly treated in the context of repeated games. It is plausible that, although the split or steal game played on Golden Balls represents a one-time interaction between the contestants, owing to the televised nature of the show the contestants may still view their interaction on the show as one of many repeated games that they will play with various other opponents, including potential customers, co-workers, and employers. We note, however, that whatever the source behind people's aversion towards lying may be, malleability of statements to ex-post justification appears to be a reasonable criterion for judging the signals' truthfulness.

Second, the televised nature of the show inevitably raises the question of whether costs of lying may be less important in the untelevized, everyday interactions between people. Previous research has shown indeed that people behave differently under public scrutiny than they would in more anonymous situations, although in the context of Golden Balls it is hard to say a priori how the fact that the show is aired on TV affects behavior. One might argue that contestants may take special care to avoid being unscrupulous liars, but one can also argue that a contestant would not want to be seen as someone who cannot play a game in which winning requires skillful lying. With these concerns in mind, our conclusion is that when people do feel scruples about lying malleability may still function as a criterion for judging the averseness of the lie.

Finally, our finding that unmalleable statements can credibly predict future action naturally raises the question of whether people also recognize these as cues to assess the opponent's intentions. Recently, He, Offerman, and van de Ven (2016) reported that communication enhances cooperation by virtue of lie averse individuals keeping their promises to cooperate but more importantly by allowing conditionally cooperative individuals to accurately assess the opponent's type; i.e., distinguish the cooperative from the uncooperative type. However, their data do not reveal the extent to which the subjects relied on the opponent's verbal signals in forming their beliefs. In their experiment, subjects believed that the opponent was likely to cooperate if both the subject and the opponent had made promises to each other, but this result may be because beliefs about the opponent were shaped by the subjects' own intentions which were in turn affected by their own (rather than the opponent's) promises. Unfortunately, our data on the behavior of Golden Balls contestants do not permit us to investigate this issue fully, the main complication being that we do not observe the contestants' beliefs, only their decisions to split or steal. We find that contestants' decisions are unrelated to the opponents' signals. Without accounting for contestants' beliefs, however, we cannot make conclusions on whether this no-relation is due to the contestants' inability to identify the correct cues in forming their beliefs about the opponent or to the imperfect relation between contestants' cooperative behavior and their beliefs. Our data, however, do show interactions between the contestant's own signals and those of the opponent, although with mixed results. Future research is needed to disentangle the effects of others' cheap-talk messages on people's beliefs versus behavior.

Chapter 5

A Cautionary Tale about Nudges

with SUSANNE NECKERMANN, DENNIE VAN DOLDER, AND TONG WANG

5.1 Introduction

This study seeks to examine potential low-cost verbal nudges to increase student response rates to course evaluation surveys that are conducted online. The treatments varied the text of the email messages sent to students. The nudges used in our experiment are: *impact information*, where students were told information about the impact of their evaluations; *peer information*, where they were informed that many of their peers participate; *commitment*, where students were asked whether they would participate.

Our *impact information* treatment is designed to provide meaning to the course evaluation; students were informed that their feedback would be used to help improve teaching and award good lecturers. Meaning as an incentive to induce behavior, though sparsely studied within economics, has been found to lower reservation wage and increase labor supply in laboratory settings (Ariely, Kamenica, and Prelec 2008), increase the performance of fundraising callers (Grant 2008), and increase performance of students on data-entry jobs (Kosfeld, Neckermann, and Yang 2014). Our *peer information* treatment points out the fact that in some courses, over 80% of the students gave their feedback. Providing descriptive norms has been found to influence a range of behaviors in domains including cooperation in laboratory social dilemma games (Parks, Sanna, and Berel 2001; Von Borgstede, Dahlstrand, and Biel 1999), substance abuse among college students (Haines and Spear 1996; Perkins 2003), environmental protection and energy conservation (Allcott 2011; Ayres, Raseman, and Shih 2012; Cialdini, Reno, and Kallgren 1990; Goldstein, Cialdini, and Griskevicius, 2008; Schultz et al. 2007), and political elections (Gerber and Rogers 2009). The *commitment* treatment is based on the commitment/consistency principle from psychology, which states that people are more likely to behave in ways that are congruent with a position which they previously committed themselves to (Aronson 1992; Cialdini 1993). Similar treatments in the literature include pre-election surveys (Greenwald et al. 1987; Mann 2005), though Smith, Gerber, and Orlich's (2003) large-scale replication of Greenwald et al.'s 1987 study yielded a null result.

Our field experiment included all bachelor and master students enrolled at the Erasmus School of Economics during the academic year 2013 – 2014. Surprisingly, we find opposite treatment effects for students enrolled in bachelor and master courses: while the treatments worked for master students, they actually lowered response rates for bachelor students. These different effects can partly be explained by observable differences between the two groups, but not entirely: the *impact information* treatment has opposing effects for the two groups even after accounting for observable differences.

These results add to the extensive literature on the power of influence tactics to influence behavior. Moreover, we contribute to the increasing interest in understanding heterogeneities of treatment effects to better target particular groups for nudges.[23] Cialdini et al. (1999), in an experiment involving university students in Poland and the United States, point out that the commitment/consistency principle had greater impact on Americans, whereas the social proof principle had greater impact on Poles. Chong et al. (2013) test a variety of treatments to induce recycling in Peru. These messages include those similar to our *impact information* and *peer information* treatments, and have been found to be effective in developed country contexts. They instead find null effects of all verbal messaging treatments in Peru. Costa and Kahn (2013) find that providing peer information in the form of a "home electricity report" aimed at increasing energy conservation works better with ideological "liberals" than "conservatives".

Our finding of opposite treatment effects for bachelor and master students calls for careful piloting of influence tactics before they are executed on a larger scale. Previous studies have stressed different effect sizes and potential null results for particular groups. In contrast, we find that treatments that are often found to be beneficial can actually be harmful for parts of the population. Hence, the worst-case scenario is not that these treatments do not help: they can actually hurt outcomes.

The opposing effects for bachelor and master students, even after accounting for these differences in observables, is especially surprising given that bachelor and master students

[23] Here, by heterogeneous treatment effects, we do not mean the "boomerang effect" studied in energy saving by Allcott (2011), Brent et al. (2015), Ferraro and Miranda (2013), Ferraro and Price (2013), and Schultz et al. (2007), and in retirement saving by Beshears et al. (2015). Their classification of groups is based on the level of target behavior before the treatment. For example, it was typically found that high-energy users are more responsive to energy saving peer information than low energy users.

are demographically highly similar. Unlike Cialdini et al. (1999) where Americans and Poles differ in individualistic-collectivistic orientation, Chong et al. (2013) where the American population and Peruvian population can be argued to be different in long-term pro-recycling views, and Costa and Kahn (2013) where liberals and conservatives differ in their energy consciousness, there is no a priori reason to expect bachelor and master students to respond differently to our treatments. Overall, our results thus suggest that nudges are more unpredictable than previously thought.

5.2 Experimental design and data

The experiment took place in the academic year 2013 – 2014 at the Erasmus School of Economics (ESE). The academic year at ESE is organized into five blocks. Courses begin and end within a block. At the end of the block, students can evaluate the quality of courses attended by filling out an online survey. Because participation in the evaluation surveys is voluntary, participation rates are generally low, averaging at 24% for the ten blocks (two academic years) that preceded the year in which our experiment took place.

The evaluation surveys are administered through ESE's online platform for communication and course registration. Typically, one week before the end of a block students receive an email message announcing the opening of the surveys, as well as giving instructions about where to find them. The surveys remain open for three weeks in total. Students receive a second email message reminding them to fill out the surveys one week before the surveys close.

Our experimental treatments varied the text of the email messages sent to students, leaving the timing and the frequency of the emails unchanged. The email message in the *control* treatment contained a basic text informing students about the survey opening and closing dates, as well as where to find the surveys. In addition to the control treatments, three different treatments were designed each in turn to test a particular persuasion tactic hypothesized to increase participation. All treatment emails are reproduced in Appendix 5.A (see Figures 5.A1-A5).

The first treatment was designed to test whether students' participation is sensitive to their perception of usefulness of their evaluations. In this *impact information* treatment, the email message contained, in addition to the basic message about when and how the surveys can be filled out, information about the impact of students' evaluations. Specifically, the email message informed students that their evaluations are used to reward lecturers with good evaluation scores and pressure lecturers with poor evaluation scores to improve their teaching quality.

Based on the findings in the literature about the effectiveness of peer information in inducing pro-social behavior, the second treatment was designed to test whether students increase their participation if they learn that many of their peers participate. Because average participation rate has generally been low, the information about high participation rates of peers provided in this *peer information* treatment was based on participation rates for courses with the highest rates. Students were told that over 80% of the participants provided feedback in some courses.

The third treatment was designed to test the effectiveness of the well-known *commitment* tactic. Students receive two emails, one at the opening and a second a week before the closing of the evaluation surveys. In the *control*, *impact information*, and *peer information* treatments, students received two identical email messages. In the *commitment* treatment, the first email prompted students to commit to participating by asking them a yes-or-no question about whether they intended to fill out the surveys. The second email contained the same basic message as in the control treatment.

Treatment assignment occurred at the beginning of the academic year, and remained unchanged throughout the year (for all five blocks). All students registered on ESE's online platform at the treatment assignment time were randomly assigned to one of the four—*control*, *impact information*, *peer information*, or *commitment*—treatments. Because registration on ESE's online platform is automatic as soon as a student is issued a student ID number, and de-registration requires the student's explicit request, the list of students registered on the platform—and therefore, assigned to a treatment—also included past students who no longer attended courses at ESE but had not requested to be de-registered. The survey participation rate for a course is defined based on the actual number of students who requested an exam for that course.

Table 5.1 shows the number of students assigned to each treatment for bachelor and master courses respectively. Although the treatments assigned to students remained unchanged throughout the five blocks of the academic year, the number of students per treatment varied across blocks because the number of courses that students attended varied across blocks. Thus for example, if a student assigned to the control treatment attended courses in blocks 1 – 3 but did not attend courses in blocks 4 and 5, the number of students assigned to the control treatment would be one less in blocks 4 and 5.

TABLE 5.1 Number of students and observations per treatment

	Total	Control	Impact Information	Peer Information	Commitment			
					total	"no"	"yes"	no response
				Bachelor Level				
Number of Students								
academic year	3,485	889	884	842	870	33	169	841
block 1	2,847	728	719	692	708	15	101	592
block 2	2,971	751	758	716	746	11	83	652
block 3	2,773	712	695	680	686	6	62	618
block 4	2,589	654	666	634	635	7	54	574
block 5	2,411	610	608	602	591	1	35	555
Number of Observations								
academic year	28,029	7,209	7,043	6,904	6,873	86	725	6,062
block 1	5,945	1,545	1,499	1,443	1,458	32	218	1,208
block 2	6,151	1,570	1,531	1,532	1,518	20	178	1,320
block 3	6,532	1,690	1,617	1,630	1,595	16	144	1,435
block 4	4,908	1,262	1,255	1,192	1,199	15	116	1,068
block 5	4,493	1,142	1,141	1,107	1,103	3	69	1,031
				Master Level				
Number of Students								
academic year	1,552	382	374	405	391	7	85	381
block 1	1,341	327	312	362	340	5	56	279
block 2	1,249	319	298	318	314	2	37	275
block 3	1,013	256	237	262	258	0	27	231
block 4	983	247	224	258	254	0	19	235
block 5	121	30	19	37	35	1	1	33
Number of Observations								
academic year	8,434	2,138	1,960	2,214	2,122	18	282	1,822
block 1	3,571	910	819	956	886	14	155	717
block 2	1,954	497	478	507	472	3	56	413
block 3	1,272	328	293	322	329	0	39	290
block 4	1,516	373	351	392	400	0	31	369
block 5	121	30	19	37	35	1	1	33

Our outcome variable of interest is a student's decision to participate or not in the evaluation survey for each attended course. Thus, the number of observations (also shown in Table 5.1) represents the number of student-course decisions to fill out the evaluation survey or not.

5.3 Results

Table 5.2 shows the rates of students' participation in the evaluation surveys across the four treatments for bachelor and master courses respectively. Overall, the treatment effects on students attending master courses were in the opposite direction from those attending bachelor courses. All three treatments backfired for bachelor courses. The *impact information* had the biggest effect, consistently lowering the evaluation response rates for the first 4 blocks. These same treatments, however, increased evaluation response rates in the master courses.

TABLE 5.2 Response rate overview

	Total	Control	Impact Information	Peer Information	Commitment			
					total	"no"	"yes"	no response
Bachelor Level								
academic year	18.76%	20.39%	16.84%***	19.16%**	18.62%***	34.88%	79.59%	11.10%
block 1	21.65%	24.14%	20.48%***	21.97%*	19.89%***	18.75%	68.35%	11.18%
block 2	17.43%	19.81%	15.22%***	18.34%	16.27%***	35.00%	75.28%	8.03%
block 3	19.18%	20.59%	15.83%***	20.12%	20.13%	68.75%	91.67%	12.40%
block 4	21.37%	23.22%	19.04%***	21.06%*	22.19%	40.00%	87.93%	14.79%
block 5	13.31%	12.70%	13.23%	13.19%	14.14%	0.00%	86.96%	9.31%
Master Level								
academic year	27.45%	24.84%	28.98%^^^	28.73%^^^	27.33%^^	16.67%	81.21%	19.10%
block 1	27.72%	23.74%	29.91%^^^	29.08%^^^	28.33%^^	14.29%	76.77%	18.13%
block 2	25.54%	26.56%	25.10%	27.42%	22.88%*	0.00%	80.36%	15.25%
block 3	31.76%	26.83%	34.81%^^	31.06%	34.65%^^	-	89.74%	27.24%
block 4	26.39%	24.40%	27.64%	28.83%^	24.75%	-	93.55%	18.97%
block 5	18.18%	13.33%	21.05%	16.22%	22.86%	100.00%	100.00%	18.18%

NOTES: Proportion tests are used to compare response rates of treatments to *control*.
***: significantly lower than *control* treatment at 1% level.
**: significantly lower than *control* treatment at 5% level.
*: significantly lower than *control* treatment at 10% level.
^^^: significantly higher than *control* treatment at 1% level.
^^: significantly higher than *control* treatment at 5% level.
^: significantly higher than *control* treatment at 10% level.

These opposite effects for bachelor and master courses are puzzling. Next, we will explore three key differences between bachelor and master courses and investigate whether these differences can help explain the opposite effects.

First, since many students in the Netherlands complete their bachelor and master at the same university, they have been affiliated longer with the university. ESE's online platform allowed us to extract information on the year in which a student first began a study program at Erasmus University Rotterdam. We define a student's affiliation length as the number of years passed since the student first enrolled for a study program at Erasmus University at the time our experiment started (September 2013). Table 5.3 reports summary statistics on students' affiliation lengths. They are not significantly different among the four treatment groups, but indeed master students have significantly higher affiliation lengths.

Second, master students attend fewer courses compared to bachelor students. Table 5.4 reports summary statistics of the number of courses registered, extracted from our exam registration data. The number of courses registered mostly does not differ among the four treatment groups. Master students took fewer courses over the year and for all blocks except the first one. Finally, master courses have fewer students, as shown in Table 5.5.

All three factors could potentially influence how our treatments work. We run two logit regressions on a student's decision of whether or not to complete the evaluation of a course, to see the interaction effects of these factors with treatments. To facilitate the interpretation of average marginal effects course size is standardized at the course level, affiliation duration is standardized at the student level, and number of courses registered is standardized at the student-block level. Hence the unit of average marginal effects for these three factors is one standard deviation from the mean.

Table 5.6 column (1) reports average marginal effects of treatments for bachelor and master courses, controlling for course level and the three factors, and including block fixed effects. All reported values were estimated with robust standard errors clustered on the student-block level. Our observations from Table 2 are roughly confirmed. The *impact information* treatment reduced a bachelor student's likelihood of participating in the course evaluation. However, a master student was more likely to fill out the evaluation survey when informed about how the university uses students' course evaluations. *Peer information* treatment has a negative though insignificant effect on bachelor courses, while having large positive effect on master courses. *Commitment* treatment has negative effect on bachelor courses and positive but insignificant effect on master courses. Larger course size, longer affiliation duration, and more number of courses registered reduce the propensity to fill in an evaluation. Table 5.6 column (2), compared to column (1), allows additionally interaction effects between the three factors and treatments.

TABLE 5.3 Student characteristics: affiliation length

	Total	Control	Impact Information	Peer Information	Commitment	p
			Bachelor Level			
academic year	1.83 (2.57)	1.79 (2.40)	1.88 (2.54)	1.90 (2.84)	1.75 (2.48)	0.59
block 1	1.41 (2.14)	1.34 (1.81)	1.47 (2.25)	1.48 (2.36)	1.36 (2.13)	0.52
block 2	1.56 (2.31)	1.54 (2.13)	1.56 (2.14)	1.62 (2.69)	1.50 (2.25)	0.81
block 3	1.68 (2.37)	1.64 (2.19)	1.70 (2.33)	1.76 (2.56)	1.62 (2.41)	0.70
block 4	1.43 (1.89)	1.42 (1.97)	1.46 (1.81)	1.51 (2.07)	1.33 (1.71)	0.38
block 5	1.31 (1.75)	1.28 (1.57)	1.35 (1.79)	1.37 (2.00)	1.25 (1.61)	0.59
			Master Level			
academic year	2.93*** (3.09)	2.84*** (2.81)	2.94*** (3.29)	3.10*** (3.38)	2.81*** (2.83)	0.54
block 1	2.76*** (2.87)	2.65*** (2.72)	2.75*** (2.90)	2.93*** (2.92)	2.70*** (2.93)	0.58
block 2	2.61*** (2.77)	2.59*** (2.47)	2.72*** (3.27)	2.72*** (2.73)	2.42*** (2.57)	0.47
block 3	2.35*** (2.53)	2.27*** (2.34)	2.33*** (2.69)	2.55*** (2.74)	2.22*** (2.33)	0.47
block 4	2.41*** (2.87)	2.39*** (2.52)	2.32*** (2.77)	2.69*** (3.58)	2.21*** (2.43)	0.27
block 5	2.30*** (2.49)	2.13*** (2.57)	1.79 (2.07)	2.51*** (2.62)	2.49*** (2.55)	0.71

NOTES: This table shows the mean and standard deviation (in parentheses) of the affiliation length of each student. Affiliation length is defined as 2013 (the year our experiment started) minus the first year a student registered in a study program of the Erasmus University Rotterdam. The p-values are from one-way ANOVA comparing among the four treatment groups.
***: significantly different from bachelor level at 1% level.
**: significantly different from bachelor level at 5% level.
*: significantly different from bachelor level at 10% level.

TABLE 5.4 Student Characteristics: Number of Courses Registered

	Total	Control	Impact Information	Peer Information	Commitment	p
			Bachelor Level			
academic year	8.14 (3.99)	8.20 (4.13)	8.11 (3.98)	8.27 (3.89)	7.98 (3.93)	0.45
block 1	2.12 (0.83)	2.14 (0.88)	2.14 (0.85)	2.10 (0.82)	2.08 (0.77)	0.43
block 2	2.13 (0.80)	2.16 (0.88)	2.11 (0.74)	2.18 (0.82)	2.08 (0.73)	0.06
block 3	2.37 (0.94)	2.39 (0.90)	2.34 (0.93)	2.41 (0.96)	2.34 (0.97)	0.38
block 4	1.90 (0.77)	1.93 (0.80)	1.89 (0.78)	1.89 (0.75)	1.89 (0.74)	0.65
block 5	1.87 (0.70)	1.87 (0.70)	1.88 (0.72)	1.84 (0.70)	1.87 (0.68)	0.79
			Master Level			
academic year	5.69*** (2.87)	5.88*** (2.98)	5.60*** (2.95)	5.65*** (2.78)	5.63*** (2.78)	0.53
block 1	2.70*** (0.93)	2.82*** (0.99)	2.69*** (0.99)	2.67*** (0.83)	2.63*** (0.88)	0.05
block 2	1.77*** (1.00)	1.79*** (1.11)	1.92*** (1.04)	1.74*** (0.93)	1.66*** (0.90)	0.01
block 3	1.31*** (0.57)	1.35*** (0.59)	1.29*** (0.58)	1.28*** (0.53)	1.34*** (0.60)	0.40
block 4	1.56*** (0.57)	1.52*** (0.55)	1.60*** (0.61)	1.53*** (0.57)	1.57*** (0.58)	0.40
block 5	1.05*** (0.25)	1.03*** (0.18)	1.00*** (0.00)	1.05*** (0.33)	1.09*** (0.28)	0.67

NOTES: This table shows the mean and standard deviation (in parentheses) of the number of courses registered for each student. The p-values are from one-way ANOVA comparing among the four treatment groups.
***: significantly different from bachelor level at 1% level.
**: significantly different from bachelor level at 5% level.
*: significantly different from bachelor level at 10% level.

Master courses have on average fewer students than bachelor courses. It is possible that free-riding incentives are stronger for students attending large courses. More importantly, the strength of free-riding incentives may have interacted with our treatments, in that our treatments may have been more effective in inducing pro-social behavior in an environment where people already felt some degree of responsibility for contributing to the public good (i.e., for students attending small courses). The hypothesis that participation likelihood may be lower when the course is large is supported by both regression (1) and (2) in Table 6. Additionally, the variable *course size* has a negative interaction effect with all treatments on likelihood of participation, especially in the *peer information* treatment where other students are explicitly mentioned.

TABLE 5.5 Course size

	Bachelor	Master
Number of Courses		
academic year	176	118
block 1	35	34
block 2	36	35
block 3	42	30
block 4	42	18
block 5	21	1
Course Size		
academic year	159.51 (174.53)	74.38*** (70.96)
block 1	169.40 (186.98)	105.03* (78.89)
block 2	171.36 (206.10)	60.57*** (48.21)
block 3	156.52 (151.65)	46.10*** (40.31)
block 4	116.86 (146.83)	87.89 (105.60)
block 5	213.95 (184.53)	121.00 (0.00)

NOTES: This table shows the number of bachelor and master level courses, and compares the mean and standard deviation (in parentheses) of course sizes.
***: significantly different from bachelor level at 1% level.
**: significantly different from bachelor level at 5% level.
*: significantly different from bachelor level at 10% level.

A sizeable fraction of the students enrolled in master courses had previously completed their bachelor degrees at the Erasmus University. As a result, master students were affiliated with the university longer than bachelor students on average. Observed from the regressions, the variable *affiliation duration* had a negative effect on likelihood of participation. Moreover, affiliation duration interacted negatively with treatments, meaning that our treatments were less effective, or more destructive, in affecting behavior of students who had been affiliated longer with the university. This is especially the case for the *commitment* treatment. Thus, controlling for this interaction does not help explain the divide in the treatment effects for master and bachelor students.

TABLE 5.6 Regression analysis: Bachelor vs. Master

	Dependent variable: *complete* logit regression AMEs in percentage points (std.error) (1)	 (2)
Treatments (Bachelor)		
impact information	-3.80*** (0.91)	-3.60*** (0.91)
peer information	-1.43 (0.94)	-0.94 (0.96)
commitment	-1.97** (0.93)	-1.51 (0.95)
Treatments (Master)		
impact information	3.84** (1.84)	3.37* (2.01)
peer information	4.13** (1.76)	1.91 (1.93)
commitment	1.83 (1.74)	0.27 (1.91)
Course size * Treatment		
impact information		-1.21*** (0.41)
peer information		-3.00*** (0.44)
commitment		-2.79*** (0.40)
Affiliation duration * Treatment		
impact information		-5.78*** (1.23)
peer information		-6.92*** (1.66)
commitment		-9.14*** (1.26)
Number of courses registered * Treatment		
impact information		-2.29*** (0.53)
peer information		-0.56 (0.58)
commitment		-1.61*** (0.56)
Control variables		
master	6.54*** (0.80)	6.56*** (0.81)
course size	-2.05*** (0.21)	-2.08*** (0.21)
affiliation duration	-7.57*** (0.68)	-7.62*** (0.67)
number of courses registered	-1.62*** (0.28)	-1.62*** (0.28)
Block fixed effect	included	included
Clustering	id-block	id-block
Number of observations	36,463	36,463
Pseudo R-squared	0.0305	0.0321

***: significant at 1% level.
**: significant at 5% level.
*: significant at 10% level.

Master students are registered for fewer courses than bachelor students in each block except for block 1. Having fewer surveys to fill in could make our treatments more effective. This is indeed the case for *impact information* and *commitment* treatment, as shown by the interaction terms of the number of courses registered and treatments in regression (2) of Table 5.6.

Looking at the treatment effects after accounting for observed differences between bachelor and master students, the *peer information* and *commitment* treatments have no significant main effects, while the effects of *impact information* remains significant, albeit smaller. Part of the difference between bachelor and master students can thus be accounted for by these observables, but a part of the difference remains unexplained.

5.4 Conclusion

We conduct a field experiment testing the efficacy of three low-cost, verbal persuasion tactics in improving education evaluation survey response rate. The experiment included all bachelor and master students enrolled at the Erasmus School of Economics during the academic year 2013 – 2014. The treatments varied the text of the email messages sent to students. In addition to the practical information included in the control group, *impact information* treatment included information about the impact of students' evaluations; *peer information* informs students that many of their peers participate; *commitment* treatment asked students whether they would participate.

We find opposing effects of our treatments for bachelor and master students. While overall the treatments worked for master students, they hurt bachelor response rates. Three key observable differences between bachelor and master students that may explain this difference are explored: at the student level, affiliation duration with the university and number of courses registered that block, and at the course level, course size. We find that course size and number of courses registered can explain the bachelor-master pattern, while affiliation duration cannot. Even after controlling for these observable differences, the *impact information* treatment still influences behavior in the opposite direction for bachelor and master students.

Taken together, our results show the importance to conducted randomized control trails before applying these verbal persuasion tactics in a particular setting. In our experiment, students attending master courses were more likely to participate in their course evaluations when they were informed of the impact that these evaluations can have on teaching, or when told that many of their peers also complete these surveys, or being asked if they would complete these surveys beforehand. Somewhat disturbing was the negative effects that these same treatments had on bachelor students' participation. Thus, our costless treatments had costly unintended effects on bachelor students.

5.A Appendix

FIGURE 5.A1 Email message in *control* treatment

FOR ENGLISH SCROLL DOWN

Beste student,

Je kunt de kwaliteit van jouw vakken evalueren door met je ERNA gebruikersnaam en wachtwoord in te loggen op SIN-Online. Klik op *My Page*, scroll omlaag en klik op *Questionnaires Waiting*.

Het invullen van de vragenlijst zal slechts 5 minuten in beslag nemen. De vragenlijst is beschikbaar tot een week na het tentamen.

Bij voorbaat dank voor je deelname,

Ria Koolen, Onderwijs Service Centrum, ESE

Dear student,

You can evaluate the quality of your courses by logging into SIN-Online using your ERNA username and password. Click on *My Page*, scroll down, then click on *Questionnaires Waiting*.

It will take you only 5 minutes to complete the survey. The survey will close one week after the examination.

Thank you in advance for your participation,

Ria Koolen, Education Service Center, ESE

FIGURE 5.A2 Email message in *impact information* treatment

FOR ENGLISH SCROLL DOWN

Beste student,

Jouw evaluatie wordt gebruikt om de onderwijskwaliteit van jouw vakken te bepalen:

1. Docenten gebruiken de resultaten als feedback om hun vakken te verbeteren.

2. Evaluatiescores worden besproken binnen de afdeling: goede scores worden erkend en beloond; slechte scores resulteren in druk van superieuren en collega's om onderwijs te verbeteren.

Jouw mening is nodig voor een accurate evaluatie van jouw vakken en docenten.

Je kunt de kwaliteit van jouw vakken evalueren door met je ERNA gebruikersnaam en wachtwoord in te loggen op SIN-Online. Klik op *My Page*, scroll omlaag en klik op *Questionnaires Waiting*.

Het invullen van de vragenlijst zal slechts 5 minuten in beslag nemen. De vragenlijst is beschikbaar tot een week na het tentamen.

Bij voorbaat dank voor je deelname,

Ria Koolen, Onderwijs Service Centrum, ESE

Dear student,

Your evaluation is used to assess the teaching quality of your courses:

1. Lecturers use the results as feedback to improve their courses.

2. Evaluation scores are discussed at the department: good scores are recognized and rewarded; poor scores result in pressure from superiors and colleagues to improve teaching.

Your opinion is needed for an accurate evaluation of your courses and lecturers.

You can evaluate the quality of your courses by logging into SIN-Online. using your ERNA username and password. Click on *My Page*, scroll down, then click on *Questionnaires Waiting*.

It will take you only 5 minutes to complete the survey. The survey will close one week after the examination.

Thank you in advance for your participation,

Ria Koolen, Education Service Center, ESE

FIGURE 5.A3 Email message in *peer information* treatment

FOR ENGLISH SCROLL DOWN

Beste student,

Studenten op de ESE nemen actief deel aan vakevaluaties. **In een aantal vakken levert meer dan 80% van de studenten feedback.**

Je kunt de kwaliteit van jouw vakken evalueren door met je ERNA gebruikersnaam en wachtwoord in te loggen op SIN-Online. Klik op *My Page*, scroll omlaag en klik op *Questionnaires Waiting*.

Het invullen van de vragenlijst zal slechts 5 minuten in beslag nemen. De vragenlijst is beschikbaar tot een week na het tentamen.

Bij voorbaat dank voor je deelname,

Ria Koolen, Onderwijs Service Centrum, ESE

Dear student,

Students at ESE actively participate in course evaluation. **In a number of courses over 80% of students provide their feedback.**

You can evaluate the quality of your courses by logging into SIN-Online using your ERNA username and password. Click on *My Page*, scroll down, then click on *Questionnaires Waiting*.

It will take you only 5 minutes to complete the survey. The survey will close one week after the examination.

Thank you in advance for your participation,

Ria Koolen, Education Service Center, ESE

FIGURE 5.A4 Email message and reminder in *commitment* treatment

FOR ENGLISH SCROLL DOWN

Beste student,

Het is voor ons nuttig om vooraf te weten hoe veel evaluatievragenlijsten zullen worden ingevuld. **Laat ons alsjeblieft weten of je van plan bent om de vakevaluatie in te vullen, door "yes" of "no" te klikken na de volgende link: Link.**

Je kunt de kwaliteit van jouw vakken evalueren door met je ERNA gebruikersnaam en wachtwoord in te loggen op SIN-Online. Klik op *My Page*, scroll omlaag en klik op *Questionnaires Waiting*.

Het invullen van de vragenlijst zal slechts 5 minuten in beslag nemen. De vragenlijst is beschikbaar tot een week na het tentamen.

Bij voorbaat dank voor je deelname,

Ria Koolen, Onderwijs Service Centrum, ESE

Dear student,

It is helpful for us to know in advance how many evaluation surveys will be completed. **Please tell us whether you intend to complete the course evaluation, by clicking "yes" or "no" using the following link: Link.**

You can evaluate the quality of your courses by logging into SIN-Online using your ERNA username and password. Click on *My Page*, scroll down, then click on *Questionnaires Waiting*.

It will take you only 5 minutes to complete the survey. The survey will close one week after the examination.

Thank you in advance for your participation,

Ria Koolen, Education Service Center, ESE

Reminder

FOR ENGLISH SCROLL DOWN

Beste student,

Je kunt de kwaliteit van jouw vakken evalueren door met je ERNA gebruikersnaam en wachtwoord in te loggen op SIN-Online. Klik op *My Page*, scroll omlaag en klik op *Questionnaires Waiting*.

Het invullen van de vragenlijst zal slechts 5 minuten in beslag nemen. De vragenlijst is beschikbaar tot een week na het tentamen.

Bij voorbaat dank voor je deelname,

Ria Koolen, Onderwijs Service Centrum, ESE

Dear student,

You can evaluate the quality of your courses by logging into SIN-Online using your ERNA username and password. Click on *My Page*, scroll down, then click on *Questionnaires Waiting*.

It will take you only 5 minutes to complete the survey. The survey will close one week after the examination.

Thank you in advance for your participation,

Ria Koolen, Education Service Center, ESE

Summary

This thesis has examined the ambiguity about others' behavior in social dilemmas, how it affects decisions, and how it can be mitigated. Three types of (stylized) social dilemma situations were studied: the trust game, a variant of the prisoner's dilemma, and a variant of the public good game. The analyses were conducted on data collected from laboratory and field experiments, as well as on field data from a TV game show.

Chapter 2 extended a method for measuring ambiguity attitudes for uncertain events in individual choice to game situations. Unlike the measurements of beliefs in experimental games made under the behaviorally invalid but widely adopted assumption that subjects are ambiguity neutral, the method of Chapter 2 can correct for ambiguity attitudes and thus give accurate measures of beliefs even if the assumption of ambiguity-neutrality is violated. Application of the method to a trust game sheds new light on the determinants of people's decisions to trust others. While previous studies of trust decisions found no role for risk attitudes, ambiguity attitudes were found to matter. Ambiguity aversion enhances distrusting decisions. Higher trust in the sense of believing more that others are trustworthy (now corrected for ambiguity attitude) enhances trusting decisions. On the basis of revealed preference data, the study in Chapter 2 also confirmed that introspective trust survey questions do capture trust in the commonly accepted sense of belief in trustworthiness of others.

Chapter 3 used the method of Chapter 2 to reexamine betrayal aversion. Betrayal aversion describes the phenomenon of people experiencing more aversion toward uncertainty about a fellow human's betrayal than they would toward uncertainty about forces of nature. In studying the nature vs. human divide in people's attitude toward ambiguity, Chapter 3 showed that previously reported betrayal aversion can be explained by ambiguity aversion. People were found to be more averse toward nature-generated ambiguity than toward betrayal ambiguity.

Chapter 4 used data from a TV game show to investigate the credibility of cheap talk in a variant of the prisoner's dilemma game. Using the free-form pre-play communication exchanged between contestants of the game show, Chapter 3 investigated whether the distinction between different types of statements added to the predictive power of the

contestants' cheap talk. Building on psychological insights, the hypothesis was proposed that contestants who wished to defect preferred to make statements that allowed them to deny the fact that they had lied to the opponent during the pre-play talk. Specifically, Chapter 3 proposed a typology of statements according to how malleable these are to ex-post reinterpretation as truths, with less malleable statements being more predictive of cooperative behavior than more malleable statements. The analysis of the data showed that malleability is indeed a plausible criterion for judging the credibility of cheap talk.

Chapter 5 tested popular nudging techniques. A field experiment was conducted among Erasmus University students, designed to increase their (voluntary) participation rate in the online evaluation surveys of their courses. The experiment randomly assigned students registered at the university to one of three different messaging nudges. The three nudge treatments emphasized to students the real impact of their course evaluations on instructor incentives, descriptive social norms of high participation rates, and commitment to fill out the evaluation survey. The results turned out to be of opposite direction for seemingly similar groups, with the Master students increasing their participation rate in response to all three nudges, and Bachelor students decreasing their participation rate as a result of nudging. These results tell a cautionary tale about the context and target dependence of nudging efficacy.

Bibliography

ABDELLAOUI, MOHAMMED, HAN BLEICHRODT, EMMANUEL KEMEL, AND OLIVIER L'HARIDON (2014) "Beliefs and Attitudes towards Natural Sources of Ambiguity," working paper.

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, 695–723.

ALLCOTT, HUNT (2011) "Social Norms and Energy Conservation" *Journal of Public Economics*, 95, 1082–1095.

ANGELOPOULOS, ANGELOS AND LEONIDAS C. KOUTSOUGERAS (2015) "Value Allocation under Ambiguity," *Economic Theory* 59, 147–167.

ANSCOMBE, FRANK J. AND ROBERT J. AUMANN (1963) "A Definition of Subjective Probability," *Annals of Mathematical Statistics* 34, 199–205.

ANTONOVICS, KATE, PETER ARCIDIACONO, AND RANDALL WALSH (2005) "Games and Discrimination Lessons From The Weakest Link," *Journal of Human Resources* 40, 918–947.

ARIELY, DAN, EMIR KAMENICA, AND DRAŽEN PRELEC (2008) "Man's Search for Meaning: The Case of Legos" *Journal of Economic Behavior and Organization* 67, 671–677.

ARMANTIER, OLIVIER AND NICOLAS TREICH (2009) "Subjective Probabilities in Games: An Application to the Overbidding Puzzle," *International Economic Review* 50, 1013–1041.

ARONSON, ELLIOT (1992) "The Return of the Repressed: Dissonance Theory Makes a Comeback" *Psychological Inquiry* 3, 303–311.

AUMANN, ROBERT J. AND DRÈZE, JACQUES H. (2009) "Assessing Strategic Risk," *American Economic Journal: Microeconomics* 1, 1–16.

AYRES, IAN, SOPHIE RASEMAN, AND ALICE SHIH (2012) "Evidence from Two Large Field Experiments that Peer Comparison Feedback can Reduce Residential Energy Usage" *Journal of Law, Economics, and Organization* 29, 992–1022.

BADE, SOPHIE (2015) "Randomization Devices and the Elicitation of Ambiguity-Averse Preferences," *Journal of Economic Theory* 159, 221–235.

BAILLON, AURÉLIEN, ZHENXING HUANG, ASLI SELIM, AND PETER P. WAKKER (2016) "Measuring Ambiguity Attitudes for All (Natural) Events," working paper.

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, 418–443.

BALLIET, DANIEL (2010) "Communication and Cooperation in Social Dilemmas: A Meta Analytic Review," *Journal of Conflict Resolution* 54, 39–57.

BATTIGALLI, PIERPAOLO, SIMONE CERREIA-VIOGLIO, FABIO MACCHERONI, AND MASSIMO MARINACCI (2015) "Self-Confirming Equilibrium and Model Uncertainty," *American Economic Review* 105, 646–677.

BATTIGALLI, PIERPAOLO, GARY CHARNESS, AND MARTIN DUFWENBERG (2013) "Deception: The Role of Guilt," *Journal of Economic Behavior and Organization* 93, 227–232.

BEETSMA, ROEL M.W.J. AND PETER C. SCHOTMAN (2001) "Measuring Risk Attitudes in a Natural Experiment: Data from the Television Game Show Lingo," *The Economic Journal* 111, 821–848.

BELLEMARE, CHARLES AND SABINE KRÖGER (2007) "On Representative Social Capital," *European Economic Review* 51, 183–202.

BELLEMARE, CHARLES, SABINE KRÖGER, AND ARTHUR VAN SOEST (2008) "Measuring Inequity Aversion in a Heterogeneous Population Using Experimental Decisions and Subjective Probabilities," *Econometrica* 76, 815–839.

BELOT, MICHÈLE, V. BHASKAR, AND JEROEN VAN DE VEN (2010) "Promises and Cooperation: Evidence from a TV Game Show," *Journal of Economic Behavior and Organization* 73, 396–405.

BENNETT, RANDALL W. AND KENT A. HICKMAN (1993) "Rationality and the 'Price is Right'," *Journal of Economic Behavior and Organization* 21, 99–105.

BERG, JOYCE, JOHN DICKHAUT, AND KEVIN MCCABE (1995) "Trust, Reciprocity, and Social History," *Games and Economic Behavior* 10, 122–142.

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, 954–970.

BESHEARS, JOHN, JAMES J. CHOI, DAVID LAIBSON, BRIGITTE C. MADRIAN, AND KATHERINE L. MILKMAN (2015) "The Effect of Providing Peer Information on Retirement Savings Decisions" *Journal of Finance* 70, 1161–1201.

BLANCO, MARIANA, DIRK ENGELMANN, ALEXANDER K. KOCH, AND HANS-THEO NORMANN (2010) “Belief Elicitation in Experiments: Is there a Hedging Problem?,” *Experimental Economics* 13, 412–438.

BOHNET, IRIS AND RICHARD ZECKHAUSER (2004) “Trust, Risk and Betrayal,” *Journal of Economic Behavior and Organization* 55, 467–484.

BOHNET, IRIS, FIONA GREIG, BENEDIKT HERRMANN, AND RICHARD ZECKHAUSER (2008) “Betrayal Aversion: Evidence from Brazil, China, Oman, Switzerland, Turkey, and the United States,” *American Economic Review* 98, 294–310.

BOMMIER, ANTOINE (2014) “A Dual Approach to Ambiguity Aversion,” working paper.

BRENT, DANIEL A., JOSEPH H. COOK, AND SKYLAR OLSEN (2015) “Social Comparisons, Household Water Use, and Participation in Utility Conservation Programs: Evidence from Three Randomized Trials” *Journal of the Association of Environmental and Resource Economists* 2, 597–627.

BURKS, ARTHUR W. (1977) “*Chance, Cause, Reason (An Inquiry into the Nature of Scientific Evidence)*.” The University of Chicago Press, Chicago.

CAMERER, COLIN (2003) “*Behavioral Game Theory: Experiments in Strategic Interaction*.” Princeton University Press, New Jersey.

CAMERER, COLIN F. AND RISTO KARJALAINEN (1994) “Ambiguity-Aversion and Non-Additive Beliefs in Non-Cooperative Games: Experimental Evidence,” *Models and Experiments in Risk and Rationality*, Springer, Netherlands, 325–358.

CAPPELEN, ALEXANDER W., ERIK O. SORENSEN, AND BERTIL TUNGODDEN (2013) “When Do We Lie?,” *Journal of Economic Behavior and Organization* 93, 258–265.

CHAKRAVARTY, SURAJEET AND DAVID KELSEY (2016) “Ambiguity and Accident Law,” *Journal of Public Economic Theory* 18.

CHARK, ROBIN AND SOO HONG CHEW (2015) “A Neuroimaging Study of Preference for Strategic Uncertainty,” *Journal of Risk and Uncertainty* 50, 209–227.

CHARNESS, GARY AND MARTIN DUFWENBERG (2006) “Promises and Partnership,” *Econometrica* 74, 1579–1601.

CHARNESS, GARY AND MARTIN DUFWENBERG (2010) “Bare Promises: An Experiment,” *Economics Letters* 107, 281–283.

CHERIDITO, PATRICK, FREDDY DELBAEN, SAMUEL DRAPEAU, AND MICHAEL KUPPER (2015) “Stochastic Order-Monotone Uncertainty-Averse Preferences,” HTTP://SSRN.COM/ABSTRACT=2572745.

CHEW, SOO HONG, RICHARD P. EBSTEIN, AND SONGFA ZHONG (2012) "Ambiguity Aversion and Familiarity Bias: Evidence from Behavioral and Gene Association Studies," *Journal of Risk and Uncertainty* 44, 1–18.

CHEW, SOO HONG AND JACOB S. SAGI (2008) "Small Worlds: Modeling Attitudes toward Sources of Uncertainty," *Journal of Economic Theory* 139, 1–24.

CHONG, ALBERTO, DEAN KARLAN, JEREMY SHAPIRO, AND JONATHAN ZINMAN (2013) "(Ineffective) Messages to Encourage Recycling: Evidence from a Randomized Evaluation in Peru" *The World Bank Economic Review* 29, 180–206.

CIALDINI, ROBERT B. (1993) "*Influence: The Psychology of Persuasion*." William Morrow and Company, Inc., New York.

CIALDINI, ROBERT B., RAYMOND R. RENO, AND CARL A. KALLGREN (1990) "A Focus Theory of Normative Conduct: Recycling the Concept of Norms to Reduce Littering in Public Places," *Journal of Personality and Social Psychology* 58, 1015–1026.

CIALDINI, ROBERT B., WILHELMINA WOSINSKA, DANIEL W. BARRETT, JONATHAN BUTNER, AND MALGORZATA GORNIK-DUROSE (1999) "Compliance with a Request in Two Cultures: The Differential Influence of Social Proof and Commitment/Consistency on Collectivists and Individualists," *Personality and Social Psychology Bulletin* 25, 1242–1253.

COSTA, DORA L. AND MATTHEW E. KAHN (2013) "Energy Conservation "Nudges" and Environmentalist Ideology: Evidence from a Randomized Residential Electricity Field Experiment," *Journal of the European Economic Association* 11, 680–702.

COSTA-GOMES, MIGUEL AND GEORG WEIZSÄCKER (2008) "Stated Beliefs and Play in Normal-Form Games," *Review of Economic Studies* 75, 729–762.

DAI, ZHIXIN, ROBIN M. HOGARTH, AND MARIE CLAIRE VILLEVAL (2015) "Ambiguity on Audits and Cooperation in a Public Goods Game," *European Economic Review* 74, 146–162.

DAWES, ROBYN M. (1990) "False Consensus Effect," *Insights in Decision Making: A Tribute to Hillel J. Einhorn*, University of Chicago Press, Illinois, 179–199.

DE LARA RESENDE, JOSÉ G. AND GEORGE WU (2010) "Competence Effects for Choices Involving Gains and Losses," *Journal of Risk and Uncertainty* 40, 109–132.

DE MARCOA, GIUSEPPE AND MARIA ROMANIELLO (2015) "Variational Preferences and Equilibria in Games under Ambiguous Belief Correspondences," *International Journal of Approximate Reasoning* 60, 8–22.

DePaulo, Bella M., James J. Lindsay, Brian E. Malone, Laura Muhlenbruck, Kelly Charlton, and Harris Cooper (2003) "Cues to Deception," *Psychological Bulletin* 129, 74–118.

Di Mauro, Carmela and Massimo Finocchiaro Castro (2011) "Kindness, Confusion, or … Ambiguity?," *Experimental Economics* 14, 611–633.

Dimmock, Stephen G., Roy Kouwenberg, Olivia S. Mitchell, and Kim Peijnenburg (2015) "Estimating Ambiguity Preferences and Perceptions in Multiple Prior Models: Evidence from the Field," *Journal of Risk and Uncertainty* 51, 219–244.

Dimmock, Stephen G., Roy Kouwenberg, and Peter P. Wakker (2016) "Ambiguity Attitudes in a Large Representative Sample," *Management Science* 62, 1363–1380.

Dominiak, Adam and Peter Duersch (2015) "Benevolent and Malevolent Ellsberg Games," working paper.

Dow, James and Sérgio R.C. Werlang (1992) "Uncertainty Aversion, Risk Aversion and the Optimal Choice of Portfolio," *Econometrica* 60, 197–204.

Eckel, Catherine C. and Rick K. Wilson (2004) "Is Trust a Risky Decision?," *Journal of Economic Behavior and Organization* 55, 447–465.

Eichberger, Jürgen and David Kelsey (2011) "Are the Treasures of Game Theory Ambiguous?," *Economic Theory* 48, 313–393.

Eichberger, Jürgen, David Kelsey, and Burkhard C. Schipper (2008) "Granny versus Game Theorist: Ambiguity in Experimental Games," *Theory and Decision* 64, 333–362.

Ekman, Paul (2001) "*Telling Lies: Clues to Deceit in the Marketplace, Politics, and Marriage*." W. W. Norton and Co., New York.

Ellingsen, Tore and Magnus Johannesson (2004) "Promises, Threats and Fairness," *The Economic Journal* 114, 397–420.

Ellsberg, Daniel (1961) "Risk, Ambiguity and the Savage Axioms," *Quarterly Journal of Economics* 75, 643–669.

Epstein, Larry G. and Martin Schneider (2010) "Ambiguity and Asset Markets," *Annual Review of Financial Economics* 2, 315–346.

Erat, Sanjiv and Uri Gneezy (2012) "White Lies," *Management Science* 58, 723–733.

Farrel, Joseph and Matthew Rabin (1996) "Cheap Talk," *Journal of Economic Perspectives* 10, 103–118.

Fehr, Ernst (2009) "On the Economics and Biology of Trust," *Journal of the European Economic Association* 7, 235–266.

FEHR, ERNST, URS FISCHBACHER, BERNHARD VON ROSENBLADT, JÜRGEN SCHUPP, AND GERT G. WAGNER (2003) “A Nation-Wide Laboratory: Examining Trust and Trustworthiness by Integrating Behavioral Experiments into Representative Survey,” working paper.

FERRARO, PAUL J. AND JUAN J. MIRANDA (2013) “Heterogeneous Treatment Effects and Mechanisms in Information-Based Environmental Policies: Evidence from a Large-Scale Field Experiment,” *Resource and Energy Economics* 35, 356–379.

FERRARO, PAUL J. AND MICHAEL K. PRICE (2013) “Using Nonpecuniary Strategies to Influence Behavior: Evidence from a Large-Scale Field Experiment,” *Review of Economics and Statistics* 95, 64–73.

FETCHENHAUER, DETLEF AND DAVID DUNNING (2012) “Betrayal Aversion versus Principled Trustfulness: How to Explain Risk Avoidance and Risky Choices in Trust Games,” *Journal of Economic Behavior and Organization* 81, 534–541.

FOX, CRAIG R. AND MARTIN WEBER (2002) “Ambiguity Aversion, Comparative Ignorance, and Decision Context,” *Organizational Behavior and Human Decision Processes* 88, 476–498.

FRENCH, KENNETH R. AND JAMES M. POTERBA (1991) “Investor Diversification and International Equity Markets,” *American Economic Review* 81, 222–226.

GAMBETTA, DIEGO (2000) "Can We Trust Trust," *Trust: Making and Breaking Cooperative Relations*, electronic edition, University of Oxford, 213–237.

GERBER, ALAN S. AND TODD ROGERS (2009) “Descriptive Social Norms and Motivation to Vote: Everybody's Voting and So Should You,” *Journal of Politics* 71, 178–191.

GERTNER, ROBERT (1993) “Game Shows and Economic Behavior: Risk-taking on “Card Sharks”,” *Quarterly Journal of Economics* 108, 507–521.

GHIRARDATO, PAOLO (2002) “Revisiting Savage in a Conditional World,” *Economic Theory* 20, 83–92.

GILBOA, ITZHAK (1987) “Expected Utility with Purely Subjective Non-Additive Probabilities,” *Journal of Mathematical Economics* 16, 65–88.

GILBOA, ITZHAK AND DAVID SCHMEIDLER (1989) “Maxmin Expected Utility with a Non-Unique Prior,” *Journal of Mathematical Economics* 18, 141–153.

GILBOA, ITZHAK AND DAVID SCHMEIDLER (2003) “A Derivation of Expected Utility Maximization in the Context of a Game,” *Games and Economic Behavior* 44, 184–194.

GILBOA, ITZHAK, FABIO MACCHERONI, MASSIMO MARINACCI, AND DAVID SCHMEIDLER (2010) "Objective and Subjective Rationality in a Multiple Prior Model," *Econometrica* 78, 755–770.

GLAESER, EDWARD L., DAVID I. LAIBSON, JOSE A. SCHEINKMAN, AND CHRISTINE L. SOUTTER (2000) "Measuring Trust," *Quarterly Journal of Economics* 115, 811–846.

GNEEZY, URI (2005) "Deception: The Role of Consequences," *American Economic Review* 95, 384–394.

GOLDSTEIN, NOAH J., ROBERT B. CIALDINI, AND VLADAS GRISKEVICIUS (2008) "A Room with a Viewpoint: Using Social Norms to Motivate Environmental Conservation in Hotels," *Journal of Consumer Research* 35, 472–482.

GONZALEZ, RICHARD AND GEORGE WU (1999) "On the Shape of the Probability Weighting Function," *Cognitive Psychology* 38, 129–166.

GRANT, ADAM M. (2008) "Employees without a Cause: The Motivational Effects of Prosocial Impact in Public Service," *International Public Management Journal* 11, 48–66.

GRANT, SIMON, IDIONE MENEGHEL, AND RABEE TOURKY (2016) "Savage Games," *Theoretical Economics* 11, 641–682.

GREENWALD, ANTHONY G., CATHERINE G. CARNOT, REBECCA BEACH, AND BARBARA YOUNG (1987) "Increasing Voting Behavior by Asking People if They Expect to Vote" *Journal of Applied Psychology* 72, 315–318.

HAINES, MICHAEL AND SHERILYNN F. SPEAR (1996) "Changing the Perception of the Norm: A Strategy to Decrease Binge Drinking among College Students" *Journal of American College Health* 45, 134–140.

HAMMOND, PETER J. (1988) "Consequentialist Foundations for Expected Utility," *Theory and Decision* 25, 25–78.

HARSANYI, JOHN C. (1982) "Subjective Probability and the Theory of Games: Comments on Kadane and Larkey's Paper," *Management Science* 28, 120–125.

HARTLEY, ROGER, GAUTHIER LANOT, AND IAN WALKER (2014) "Who Really Wants to be a Millionaire? Estimates of Risk Aversion from Game Show Data," *Journal of Applied Econometrics* 29, 861–879.

HE, SIMIN, THEO OFFERMAN, AND JEROEN VAN DE VEN (2016) "The Sources of the Communication Gap," *Management Science* (forthcoming).

HEATH, CHIP AND AMOS TVERSKY (1991) "Preference and Belief: Ambiguity and Competence in Choice under Uncertainty," *Journal of Risk and Uncertainty* 4, 5–28.

HEINEMANN, FRANK, ROSEMARIE NAGEL, AND PETER OCKENFELS (2009A) "Measuring Strategic Uncertainty in Coordination Games," *Review of Economic Studies* 76, 181–221.

HEINEMANN, FRANK, ROSEMARIE NAGEL, AND PETER OCKENFELS (2009B) "Measuring Strategic Uncertainty in Coordination Games," *Review of Economic Studies* 76, 181–221.

HILBIG, BENJAMIN E. AND CORINNA M. HESSLER (2013) "What Lies Beneath: How the Distance between Truth and Lie Drives Dishonesty," *Journal of Experimental Social Psychology* 49, 263–266.

HOLT, CHARLES A. (1986) "Preference Reversals and the Independence Axiom," *American Economic Review* 76, 508–513.

HOUSER, DANIEL, DANIEL SCHUNK, AND JOACHIM WINTER (2010) "Distinguishing Trust from Risk: An Anatomy of the Investment Game," *Journal of Economic Behavior and Organization* 74, 72–81.

HSU, MING, MEGHANA BHATT, RALPH ADOLPHS, DANIEL TRANEL, AND COLIN F. CAMERER (2005) "Neural Systems Responding to Degrees of Uncertainty in Human Decision Making," *Science* 310, 9 Dec., 1680–1683.

HUCK, STEFFEN AND GEORG WEIZSÄCKER (2002) "Do Players Correctly Estimate What Others Do? Evidence of Conservatism in Beliefs," *Journal of Economic Behavior and Organization* 47, 71–85.

IVANOV, ASEN (2011) "Attitudes to Ambiguity in One-Shot Normal-Form Games: An Experimental Study," *Games and Economic Behavior* 71, 366–394.

JOHNSON, CATHLEEN, AURÉLIEN BAILLON, HAN BLEICHRODT, ZHIHUA LI, DENNIE VAN DOLDER, AND PETER P. WAKKER (2015) "Prince: An Improved Method for Measuring Incentivized Preferences," working paper.

JOHNSON, NOEL D. AND ALEXANDRA A. MISLIN (2011) "Trust Games: A Meta-Analysis," *Journal of Economic Psychology* 32, 865–889.

KARNI, EDI AND ZVI SAFRA (1987) "Preference Reversal and the Observability of Preferences by Experimental Methods," *Econometrica* 55, 675–685.

KELLNER, CHRISTIAN (2015) "Tournaments as a Response to Ambiguity Aversion in Incentive Contracts," *Journal of Economic Theory* 159, 627–655.

KELSEY, DAVID AND SARA LE ROUX (2015) "An Experimental Study on the Effect of Ambiguity in a Coordination Game," *Theory and Decision* 79, 667–688.

KELSEY, DAVID AND SARA LE ROUX (2016) "Dragon Slaying with Ambiguity: Theory and Experiments," *Journal of Public Economic Theory* 18.

KEYNES, JOHN MAYNARD (1921) "*A Treatise on Probability*." McMillan, London. 2nd edn. 1948.

KNIGHT, FRANK H. (1921) "*Risk, Uncertainty, and Profit*." Houghton Mifflin, New York.

KOSFELD, MICHAEL, SUSANNE NECKERMANN, AND XIAOLAN YANG (2014) "Knowing that You Matter, Matters! The Interplay of Meaning, Monetary Incentives, and Worker Recognition," workıng paper.

KRUPKA, ERIN L., STEPHEN LEIDER, AND MING JIANG (2016) "A Meeting of the Minds: Informal Agreements and Social Norms," *Management Science* (forthcoming).

KUGLER, TAMAR, TERRY CONNOLLY, AND LISA D. ORDÓÑEZ (2012) "Emotion, Decision, and Risk: Betting on Gambles versus Betting on People," *Journal of Behavioral Decision Making* 25, 123–134.

LAZZARINI, SERGIO G., REGINA MADALOZZO, RINALDO ARTES, AND JOSÉ DE OLIVEIRA SIQUEIRA (2005) "Measuring Trust: An Experiment in Brazil," *Brazilian Journal of Applied Economics* 9, 153–169.

LEVITT, STEVEN D. (2004) "Why are Gambling Markets Organised so Differently from Financial Markets?," *The Economic Journal* 114, 223–246.

LI, CHEN, UYANGA TURMUNKH, AND PETER P. WAKKER (2016) "Social and Strategic Ambiguity: Betrayal Aversion Revisited," working paper.

LI, CHEN, UYANGA TURMUNKH, AND PETER P. WAKKER (2016) "Ambiguity and Trust in Games," working paper.

LIST, JOHN A. (2004) "Young, Selfish and Male: Field Evidence of Social Preferences," *The Economic Journal* 114, 121–149.

LUCE, R. DUNCAN AND HOWARD RAIFFA (1957) "*Games and Decisions*." Wiley, New York.

LUNDQUIST, TOBIAS, TORE ELLINGSEN, ERIK GRIBBE, AND MAGNUS JOHANNESSON (2009) "The Aversion to Lying," *Journal of Economic Behavior and Organization* 70, 81–92.

MACHINA, MARK J. (1989) "Dynamic Consistency and Non-Expected Utility Models of Choice under Uncertainty," *Journal of Economic Literature* 27, 1622–1688.

MACHINA, MARK J. (2014) "Ambiguity Aversion with Three or More Outcomes," *American Economic Review* 104, 3814–3840.

MANN, CHRISTOPHER B. (2005) "Unintentional Voter Mobilization: Does Participation in Preelection Surveys Increase Voter Turnout?," *Annals of the American Academy of Political and Social Science*, 601, 155–168.

MAZAR, NINA, ON AMIR, AND DAN ARIELY (2008) "The Dishonesty of Honest People: A Theory of Self-Concept Maintenance," *Journal of Marketing Research* 45, 633–644.

METRICK, ANDREW (1995) "A Natural Experiment in "Jeopardy!"," *American Economic Review* (1995): 240-253.

NERI, CLAUDIA (2015) "Eliciting Beliefs in Continuous-Choice Games: A Double Auction Experiment," *Experimental Economics* 18, 569–608.

NYARKO, YAW AND ANDREW SCHOTTER (2002) "An Experimental Study of Belief Learning Using Elicited Beliefs," *Econometrica* 70, 971–1005.

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, 179–187.

OECHSSLER, JÖRG AND ALEX ROOMETS (2015) "A Test of Mechanical Ambiguity," *Journal of Economic Behavior and Organization* 119, 153–162.

PALFREY, THOMAS R. AND STEPHANIE W. WANG (2009) "On Eliciting Beliefs in Strategic Games," *Journal of Economic Behavior and Organization* 71, 98–109.

PARKS, CRAIG D., LAWRENCE J. SANNA, AND SUSAN R. BEREL (2001) "Actions of Similar Others as Inducements to Cooperate in Social Dilemmas," *Personality and Social Psychology Bulletin* 27, 345–354.

PEREA, ANDRÉS (2012) "*Epistemic Game Theory: Reasoning and Choice*." Cambridge University Press, New York.

PEREA, ANDRÉS (2014) "Belief in the Opponents' Future Rationality," *Games and Economic Behavior* 83, 231–254.

PERKINS, H. W. (2003) "The Emergence and Evolution of the Social Norms Approach to Substance Abuse Prevention," *The Social Norms Approach to Preventing School and College Age Substance Abuse: A Handbook for Educators, Counselors, and Clinicians*, Jossey-Bass, California, 3–17.

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, 38–71.

PRELEC, DRAZEN (2004) “A Bayesian Truth Serum for Subjective Data,” *Science* 306, October 2004, 462–466.

PULFORD, BRIONY D. AND ANDREW M. COLMAN (2007) “Ambiguous Games: Evidence for Strategic Ambiguity Aversion,” *Quarterly Journal of Experimental Physiology* 60, 1083–1100.

ROSS, LEE, DAVID GREENE, AND PAMELA HOUSE (1977) “The ‘False Consensus Effect:’ An Egocentric Bias in Social Perception and Attribution Processes,” *Journal of Experimental Social Psychology* 13, 279–301.

RUBINSTEIN, ARIEL AND YUVAL SALANT (2016) “ “Isn't Everyone Like Me?”: On the Presence of Self-Similarity in Strategic Interactions,” *Judgment and Decision Making* 11, 168–173.

RUTSTRÖM, ELISABET E. AND NATHANIEL T. WILCOX (2009) “Stated Beliefs versus Inferred Beliefs: A Methodological Inquiry and Experimental Test,” *Games and Economic Behavior* 67, 616–632.

SALLY, DAVID (1995) “Conversation and Cooperation in Social Dilemmas a Meta-analysis of Experiments from 1958 to 1992,” *Rationality and Society* 7, 58–92.

SAPIENZA, PAOLA, ANNA TOLDRA-SIMATS, AND LUIGI ZINGALES (2013) “Understanding Trust,” *The Economic Journal* 123, 1313–1332.

SCHMEIDLER, DAVID (1989) “Subjective Probability and Expected Utility without Additivity,” *Econometrica* 57, 571–587.

SCHNEEWEISS, HANS (1973) “The Ellsberg Paradox from the Point of View of Game Theory,” *Inference and Decision* 1, 65–78.

SCHNEIDER, FLORIAN AND MARTIN SCHONGER (2015) “Allais at the Horse Race: Testing Models of Ambiguity Aversion,” working paper.

SCHWEITZER, MAURICE E. AND CHRISTOPHER K. HSEE (2002) “Stretching the Truth: Elastic Justification and Motivated Communication of Uncertain Information,” *Journal of Risk and Uncertainty* 25, 185–201.

SCHULTZ, P. W., JESSICA M. NOLAN, ROBERT B. CIALDINI, NOAH J. GOLDSTEIN, AND VLADAS GRISKEVICIUS (2007) “The Constructive, Destructive, and Reconstructive Power of Social Norms,” *Psychological Science* 18, 429–434.

SELTEN, REINHARD (1967) “Die Strategiemethode zur Erforschung des Eingeschränkt Rationalen Verhaltens im Rahmen eines Oligopolexperimentes,” *Beiträge zur Experimentellen Wirtschaftsforschung*, J.C.B. Mohr, Tübingen, 136–168.

SERRA-GARCIA, MARTA, ERIC VAN DAMME, AND JAN POTTERS (2011) "Hiding an Inconvenient Truth: Lies and Vagueness," *Games and Economic Behavior* 73, 244–261.

SHALVI, SHAUL, JASON DANA, MICHEL J.J. HANDGRAAF, AND CARSTEN K.W. DE DREU (2011A) "Justified Ethicality: Observing Desired Counterfactuals Modifies Ethical Perceptions and Behavior," *Organizational Behavior and Human Decision Processes* 115, 181–190.

SHALVI, SHAUL, MICHEL J.J. HANDGRAAF, AND CARSTEN K.W. DE DREU (2011B) "Ethical Manoeuvring: Why People Avoid Both Major and Minor Lies," *British Journal of Management* 22, S16–S27.

SKIADAS, COSTIS (2013) "Scale-Invariant Uncertainty-Averse Preferences and Source-Dependent Constant Relative Risk Aversion," *Theoretical Economics* 8, 59–93.

SMITH, JENNIFER K., ALAN S. GERBER, AND ANTON ORLICH (2003) "Self-prophecy Effects and Voter Turnout: An Experimental Replication" *Political Psychology* 24, 593–604.

SMITH, VERNON L. (1969) "Measuring Nonmonetary Utilities in Uncertain Choices: The Ellsberg Urn," *Quarterly Journal of Economics* 83, 324–329.

SPORER, SIEGFRIED L. AND BARBARA SCHWANDT (2007) "Moderators of Nonverbal Indicators of Deception: A Meta-Analytic Synthesis," *Psychology, Public Policy, and Law* 13, 1–34.

SPRANCA, MARK, ELISA MINSK, AND JONATHAN BARON (1991) "Omission and Commission in Judgment and Choice," *Journal of Experimental Social Psychology* 27, 76–105.

SUGDEN, ROBERT (1991) "Rational Choice: A Survey of Contributions from Economics and Philosophy," *Economic Journal* 101, 751–785.

TENORIO, RAFAEL AND TIMOTHY N. CASON (2002) "To Spin or Not to Spin? Natural and Laboratory Experiments from The Price is Right," *The Economic Journal* 112, 170–195.

TRAUTMANN, STEFAN T. AND GIJS VAN DE KUILEN (2015A) "Ambiguity Attitudes," GIDEON KEREN AND GEORGE WU (EDS.) *The Wiley Blackwell Handbook of Judgment and Decision Making*, Blackwell, Oxford, 89–116.

TRAUTMANN, STEFAN T. AND GIJS VAN DE KUILEN (2015B) "Belief Elicitation: A Horse Race among Truth Serums," *The Economic Journal* 125, 2116–2135.

TVERSKY, AMOS AND PETER P. WAKKER (1995) "Risk Attitudes and Decision Weights," *Econometrica* 63, 1255–1280.

VON BORGSTEDE, CHRIS, ULF DAHLSTRAND, AND ANDERS BIEL (1999) "From Ought to is: Moral Norms in Large-Scale Social Dilemmas," *Goteborg Psychological Reports* 29, 1–17.

VON NEUMANN, JOHN AND OSKAR MORGENSTERN (1944, 1947, 1953) "*Theory of Games and Economic Behavior*." Princeton University Press, Princeton NJ.

WAKKER, PETER P. (2010) "*Prospect Theory: For Risk and Ambiguity*." Cambridge University Press, Cambridge, UK.

VANBERG, CHRISTOPH (2008) "Why Do People Keep Their Promises? An Experimental Test of Two Explanations," *Econometrica* 76, 1467–1480.

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, 2–20.

VAN DOLDER, DENNIE, MARTIJN J. VAN DEN ASSEM, COLIN CAMERER, AND RICHARD H. THALER (2015) "Standing United or Falling Divided? High Stakes Bargaining in a TV Game Show," *American Economic Review, Papers and Proceedings* 105, 402–407.

VRIJ, ALDERT (2008) "*Detecting Lies and Deceit: Pitfalls and Opportunities*." John Wiley and Sons Ltd., New Jersey.

ZUCKERMAN, MIRON, BELLA M. DEPAULO, AND ROBERT ROSENTHAL (1981) "Verbal and Nonverbal Communication of Deception," *Advances in Experimental Social Psychology* 14, 1–59.

Samenvatting | Summary in Dutch

Dit proefschrift onderzoekt ambiguïteit over het gedrag van andere personen in sociale dilemmas, hoe het beslissingen beïnvloedt, en hoe de effecten ervan gereduceerd kunnen worden. Drie soorten (gestileerde) situaties van sociale dilemma's worden bestudeerd: het vertrouwensspel, een variant van het gevangenen dilemma, en een variant van een publiek goed spel. Analyses worden uitgevoerd en de data zijn verzameld door middel van laboratorium en veld experimenten, alsmede van een tv spelshow.

Hoofdstuk 2 breidt een methode voor het meten van ambiguïteits-attituden tov onzekere gebeurtenissen uit van individuele keuzen naar spel-situaties. In tegenstelling tot het meten van geloven in experimentele spellen onder de gedragsmatig onjuiste maar desalnietemin vrijwel altijd gemaakte veronderstelling dat subjecten ambiguïteit-neutraal zijn, kan de methode van Hoofdstuk 2 corrigeren voor niet-neutrale ambiguïteits-attituden. De methode kan dus precieze maten van geloof geven zelfs als ambiguïteits-neutraliteit geschonden is. Toepassing van de methode op een vertrouwensspel werpt nieuw licht op de factoren die bepalen of mensen besluiten een ander te vertrouwen. Terwijl eerdere studies van vertrouwensspellen vonden dat risico attitude geen rol speelt, blijken ambiguïteits attituden wel een rol te spelen. Ambiguïteits-afkeer bevordert beslissingen om niet te vertrouwen. Groter vertrouwen in de zin van een sterker geloof dat anderen te vertrouwen zijn (nu gecorrigeerd voor ambiguïteits-attituden) bevordert de beslissing om te vertrouwen. Op grond van data gebaseerd op gebleken voorkeuren bevestigt Hoofdstuk 2 dat introspectieve vragenlijsten wel degelijk vertrouwen meten in de algemeen geaccepteerde betekenis van geloof dat anderen te vertrouwen zijn.

Hoofdstuk 3 gebruikt de methode van Hoofdstuk 2 om afkeer van verraad te onderzoeken. Deze afkeer beschrijft het verschijnsel dat mensen meer afkeer van onzekerheid ervaren over verraad door een medemens dan wanneer de onzekerheid krachten in de natuur betreft. In een onderzoek naar het onderscheid tussen attituden tov ambiguïteit gegenereerd door menselijk gedrag danwel door natuurlijke krachten, toont Hoofdstuk 3 dat eerder berichte afkeer van verraad verklaard kan worden door afkeer van ambiguïteit. Het hoofdstuk vindt dat mensen meer afkerig zijn van ambiguïteit gegenereerd door de natuur dan wanneer de ambiguïteit gerelateerd is aan verraad.

Hoodstuk 4 gebruikt data van een tv spelshow om de geloofwaardigheid van vrijblijvende gesprekken in een variant van het gevangenen dilemma te onderzoeken. Gebruikmakend van voorafgaande communicatie (in vrije vorm) tussen twee deelnemers aan de spelshow onderzoekt Hoofdstuk 3 of het onderscheid tussen verschillende soorten van beweringen gemaakt door de deelnemers iets toe voegt aan de voorspellende waarde van de vrijblijvende gesprekken van de deelnemers. Gebaseerd op psychologische inzichten wordt de hypothese gesteld dat deelnemers die willen afwijken liever beweringen doen die hen in staat stellen te ontkennen dat ze hun medsspeler hebben voorgelogen gedurende het voorafgaande gesprek. In het bijzonder stelt Hoofdstuk 3 een typologie voor van beweringen aangaande de mate waarin ze flexibel zijn voor herinterpretatie achteraf, waarbij minder flexibele beweringen meer samenwerkend gedrag vooorspellen dan meer flexibele beweringen. De data-analyse toont dat flexibiliteit inderdaad een plausibel criterium is voor het beoordelen van geloofwaardigheid van vrijblijvende gesprekken.

Hoofdstuk 5 onderzoekt verschillende nudge technieken. Een veld-experiment wordt uitgevoerd onder studenten aan de Erasmus Universiteit, bedoeld om hun (vrijwillige) deelname aan het online evaluatie formulier te vergroten. Het experiment wijst willekeurig studenten toe aan een van drie verschillende nudge formuleringen. De drie formuleringen benadrukken voor studenten de werkelijke impact van hun college evaluaties op beloningen voor de docent, de descriptieve sociale normen van hoge participatie graden, en de verplichting het hele evaluatie formulier in te vullen. De resultaten blijken tegenovergesteld te zijn voor schijnbaar vergelijkbare groepen. Master studenten nemen meer deel aan het invullen van de formulieren na ieder van de drie nudges, en Bachelor studenten doen dat minder als gevolg van het nudgen. Deze resultaten manen tot voorzichtigheid over hoe nudges hun doelen bereiken in verschillende contexten.

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:

634 F. ZHOU, *Essays on Mismeasurement and Misallocation on Transition Economies*
635 P. MULLER, *Labor Market Policies and Job Search*
636 N. KETEL, *Empirical Studies in Labor and Education Economics*
637 T.E. YENILMEZ, *Three Essays in International Trade and Development*
638 L.P. DE BRUIJN, *Essays on Forecasting and Latent Values*
639 S. VRIEND, *Profiling, Auditing and Public Policy: Applications in Labor and Health Economics*
640 M.L. ERGUN, *Fat Tails in Financial Markets*
641 T. HOMAR, *Intervention in Systemic Banking Crises*
642 R. LIT, *Time Varying Parameter Models for Discrete Valued Time Series*
643 R.H. KLEIJN, *Essays on Bayesian Model Averaging using Economic Time Series*
644 S. MUNS, *Essays on Systemic Risk*
645 B.M. SADABA, *Essays on the Empirics of International Financial Markets*
646 H. KOC, *Essays on Preventive Care and Health Behaviors*
647 V.V.M. MISHEVA, *The Long Run Effects of a Bad Start*
648 W. LI, *Essays on Empirical Monetary Policy*
649 J.P. HUANG, *Topics on Social and Economic Networks*
650 K.A. RYSZKA, *Resource Extraction and the Green Paradox: Accounting for Political Economy Issues and Climate Policies in a Heterogeneous World*
651 J.R. ZWEERINK, *Retirement Decisions, Job Loss and Mortality*
652 M. K. KAGAN, *Issues in Climate Change Economics: Uncertainty, Renewable Energy Innovation and Fossil Fuel Scarcity*
653 T.V. WANG, *The Rich Domain of Decision Making Explored: The Non-Triviality of the Choosing Process*
654 D.A.R. BONAM, *The Curse of Sovereign Debt and Implications for Fiscal Policy*
655 Z. SHARIF, *Essays on Strategic Communication*
656 B. RAVESTEIJN, *Measuring the Impact of Public Policies on Socioeconomic Disparities in Health*
657 M. KOUDSTAAL, *Common Wisdom versus Facts; How Entrepreneurs Differ in Their Behavioral Traits from Others*
658 N. PETER, *Essays in Empirical Microeconomics*
659 Z. WANG, *People on the Move: Barriers of Culture, Networks, and Language*

660 Z. HUANG, *Decision Making under Uncertainty-An Investigation from Economic and Psychological Perspective*
661 J. CIZEL, *Essays in Credit Risk, Banking, and Financial Regulation*
662 I. MIKOLAJUN, *Empirical Essays in International Economics*
663 J. BAKENS, *Economic Impacts of Immigrants and Ethnic Diversity on Cities*
664 I. BARRA, *Bayesian Analysis of Latent Variable Models in Finance*
665 S. OZTURK, *Price Discovery and Liquidity in the High Frequency World*
666 J. JI, *Three Essays in Empirical Finance*
667 H. SCHMITTDIEL, *Paid to Quit, Cheat, and Confess*
668 A. DIMITROPOULOS, *Low Emission Vehicles: Consumer Demand and Fiscal Policy*
669 G.H. VAN HEUVELEN, *Export Prices, Trade Dynamics and Economic Development*
670 A. RUSECKAITE, *New Flexible Models and Design Construction Algorithms for Mixtures and Binary Dependent Variables*
671 Y. LIU, *Time-varying Correlation and Common Structures in Volatility*
672 S. HE, *Cooperation, Coordination and Competition: Theory and Experiment*
673 C.G.F. VAN DER KWAAK, *The Macroeconomics of Banking*
674 D.H.J. CHEN, *Essays on Collective Funded Pension Schemes*
675 F.J.T. SNIEKERS, *On the Functioning of Markets with Frictions*
676 F. GOMEZ MARTINEZ, *Essays in Experimental Industrial Organization: How Information and Communication affect Market Outcomes*
677 J.A. ATTEY, *Causes and Macroeconomic Consequences of Time Variations in Wage Indexation*
678 T. BOOT, *Macroeconomic Forecasting under Regime Switching, Structural Breaks and High-dimensional Data*
679 I. TIKOUDIS, *Urban Second-best Road Pricing: Spatial General Equilibrium Perspectives*
680 F.A. FELSŐ, *Empirical Studies of Consumer and Government Purchase Decisions*
681 Y. GAO, *Stability and Adaptivity: Preferences over Time and under Risk*
682 M.J. ZAMOJSKI, *Panta Rhei, Measurement and Discovery of Change in Financial Markets*
683 P.R. DENDERSKI, *Essays on Information and Heterogeneity in Macroeconomics*