Dealing with Opportunism: Experimental Investigations of Measures against Dishonest Behavior

Three Essays in Behavioral Economics

By

Tobias Beck

Dealing with Opportunism: Experimental Investigations of Measures against Dishonest Behavior

Three Essays in Behavioral Economics

Dissertation zur Erlangung des akademischen Grades

Doktor der Wirtschafts- und Sozialwissenschaften

(Dr. rer. pol.)

Submitted to the Faculty of Economics and Management
of the University of Kassel

by Tobias Beck

Cologne

Submitted: November 2020

Date of viva voce: 7 June 2021

Grade: Magna cum laude (with great honor)

Revised for publication: June 2021

Summary

This dissertation aims at investigating whether and how distinct measures that are supposed to foster honest behavior work. It consists of three separate papers, in which I analyze the effectiveness and mechanisms of such measures on the basis of several laboratory experiments. Chapter 2 is dedicated to analyzing and comparing the effectiveness of distinct measures that are supposed to prevent people from lying for financial gain. Chapter 3 introduces a novel experimental design that allows the measurement of the size of the lie, the size of mistrust, and respective first-order beliefs in a two-player relationship of strategic interaction. Chapter 4 explores the mechanisms behind the effects of an honesty oath by analyzing its impact on moral deliberation and strategic decision making.

Chapter 2, titled "*Can honesty oaths, peer interaction, or monitoring mitigate lying?*", reports on several new variants of the dice experiment of Fischbacher and Föllmi-Heusi (2013) to investigate measures to reduce lying. We find that groups of two subjects lie at least to the same extent as individuals – even in a novel treatment in which one subject is instructed (and paid independently) to monitor the other. It also seems that subjects hardly lie when lying is only beneficial to others and not to themselves, even if the involved players are in a reciprocal relationship. Finally, we find that an honesty oath performs better in terms of lie mitigation compared to the four-eyes principle and monitoring.

Chapter 3, titled "*Size matters! Lying and Mistrust in the Continuous Deception Game*", presents a new experiment to measure lying and mistrust as continuous variables on an individual level. This experiment is a sender-receiver game framed as an investment game. It features two players: firstly, an advisor with complete information (i.e., the sender) who is incentivized to lie about the true value of an optimal investment and, secondly, an investor with incomplete information (i.e., the receiver) who is incentivized to invest optimally and therefore must rely on the alleged optimum reported by the advisor. The continuous message space allows observing more differentiated behavior and therefore enables testing of more sophisticated theoretical predictions. While the senders seem to make strategic considerations about their own potential to manipulate others based on the size of the lie, the receivers appear to have some endogenous preference for trust. I also find that the size of the lie and the size of mistrust do not only matter from a strategic perspective but also have an impact on how people perceive their own behavior. My findings support the conjecture that lying costs increase with the size of the lie.

Chapter 4, titled "*Quick, intuitive truth-telling and deliberate, strategic lying under oath*", investigates how oath-taking works when lying requires strategic thinking. The analysis is based on the experimental design developed in Chapter 3 and focuses on the effects of an honesty oath on deliberation time, strategic decision making, and the temporal consistency of (dis)honest behavior. On the upside, the honesty oath seems to work by making the decision to tell the truth less deliberate and more intuitive. On the downside, I observe that people who lie under oath barely reduce the sizes of their lies. In fact, my results indicate that the act of lying becomes more deliberate and strategic due to the oath. Finally, I provide evidence that oath-taking makes both the oath-takers' lying and truth-telling behavior more consistent over time. This suggests that an honesty oath encourages the oath-takers to make one basic decision whether to commit to unconditional truth-telling or to deliberately accept being dishonest for personal financial gain.

Zusammenfassung

In dieser Dissertation werden die Wirksamkeit und die Wirkmechanismen von Maßnahmen zur Förderung ehrlichen Verhaltens in einer Reihe von Laborexperimenten untersucht. Die Arbeit ist in drei Papiere unterteilt: In Kapitel 2 wird die Wirksamkeit ausgewählter Maßnahmen, die opportunistisches Verhalten verhindern sollen, miteinander verglichen. In Kapitel 3 wird ein neues experimentelles Design zur Erfassung des individuellen Ausmaßes der Lüge und des Misstrauens vorgestellt. Kapitel 4 untersucht die Wirkmechanismen eines Schwurs auf die Ehrlichkeit mit Fokus auf die Zeit, die die schwörenden Personen zum Entscheiden brauchen, und auf ihr Ausmaß an strategischer Entscheidungsfindung.

Kapitel 2 mit dem Titel „*Can honesty oaths, peer interaction, or monitoring mitigate lying?*“ legt die Ergebnisse mehrerer Variationen des Würfelexperiments von Fischbacher and Föllmi-Heusi (2013) dar. Diese Variationen basieren auf praktischen Maßnahmen, die ehrliches Verhalten in der Wirtschaft fördern sollen. Es zeigt sich, dass Gruppen von zwei Personen nicht seltener lügen als Individuen – selbst wenn einer der beiden Spieler angewiesen (und unabhängig dafür bezahlt) wird, den anderen Spieler zu überwachen. Außerdem beobachten wir, dass die Teilnehmer kaum füreinander lügen, selbst wenn sie zueinander in einem reziproken Verhältnis stehen. Schließlich zeigen wir, dass ein Schwur auf die Ehrlichkeit effektiver darin ist, lügen zu verhindern, als das Vier-Augen-Prinzip oder Überwachung.

In Kapitel 3, welches den Titel „*Size matters! Lying and Mistrust in the Continuous Deception Game*” trägt, wird ein neues experimentelles Design eingeführt, das es ermöglicht, das Ausmaß der Lüge und des Misstrauens auf individueller Ebene zu messen. Bei diesem Experiment handelt es sich um ein Sender-Receiver-Game, das als ein Investitionsspiel präsentiert wird. Darin gibt es zwei Spieler: Erstens einen Berater mit vollständiger Information (der Sender), der einen Anreiz hat bei dem Bericht des Wertes einer optimalen Investition zu lügen, und zweitens einen Investor mit unvollständiger Information (der Receiver), der incentiviert wird, optimal zu investieren, wofür er sich auf den Bericht des Beraters verlassen muss. Es zeigt sich, dass die Berater das Ausmaß ihrer Lügen strategisch nutzen, um ihre Investoren zu manipulieren. Die Investoren scheinen hingegen eine endogene Präferenz für vertrauensvolles Verhalten zu haben. Das Ausmaß der Lüge und des Misstrauens der beiden Spieler ist hierbei nicht nur strategisch relevant, sondern beeinflusst auch, wie die Spieler ihr eigenes Verhalten im Experiment bewerten. Meine Ergebnisse sind zudem konsistent mir der Annahme, dass die moralischen Kosten des Lügens mit dem Ausmaß der Lüge zunehmen.

Kapitel 4 mit dem Titel „*Quick, intuitive truth-telling and deliberate, strategic lying under oath*“ analysiert die Wirkmechanismen eines Schwurs auf die Ehrlichkeit basierend auf dem neu entwickelten experimentellen Design aus Kapitel 3. Der Fokus liegt auf der Wirkung des Schwurs auf die Entscheidungszeit, das Ausmaß strategischer Entscheidungsfindung und die zeitliche Konsistenz im Verhalten der Spieler. Einerseits führt der Schwur dazu, dass die Entscheidung, die Wahrheit zu sagen, schneller und intuitiver getroffen wird. Andererseits scheint er das Verhalten von Lügnern durchdachter und strategischer zu machen. So nimmt das Ausmaß der Lüge von Personen, die trotz des Schwurs lügen, durch den Schwur nicht ab. Weiterhin neigen die Probanden durch den Schwur dazu, entweder durchgehend die Wahrheit zu sagen oder durchgehen zu lügen. Dies deutet darauf hin, dass ein Schwur auf die Ehrlichkeit Personen vor die grundsätzliche Entscheidung stellt, entweder bedingungslos die Wahrheit zu sagen oder sich ganz bewusst für opportunistisches Verhalten zu entscheiden.

Acknowledgements

Throughout the writing of this dissertation I have received a great deal of support and assistance.

First and foremost, I would like to thank Prof. Dr. Björn Frank for his great support as my supervisor. On the one hand, he always gave me valuable feedback on my papers and guidance on how to proceed with my dissertation. On the other hand, he gave me the liberty to pursue the research questions that were (and are) most intriguing to me. When I started my doctorate, I did not know yet whether I would stay in Germany or move back to Brazil. From the beginning, Björn trusted me to do my doctorate with him, regardless of where I would decide to live. I am very thankful that he granted me such great freedom, as this allowed me to find a productive balance between my professional and personal life during my doctorate.

I also want to thank Dr. Christoph Bühren and Dr. Elina Khachatryan for being such great colleagues – even though we had to work together over a long distance most of the time. They both are always supportive and I really enjoyed working with them on our common paper, which is part of this dissertation.

Special thanks go to the Hans Böckler Foundation for accepting me into their PhD scholarship program for talented doctoral candidates, and I wish to express my gratitude to them for providing various valuable learning experiences, giving me the opportunity to grow personally and professionally.

I am also thankful to my mother, Karen Beck, for her loving support during my doctorate and beyond. In particular, I would like to thank her for the encouragement she gave me and for her persistence in proof reading my papers with the greatest accuracy.

Last but not least, I want to thank Eva Lazarus because she was always keen to learn as much as possible about my research, which motivated me beyond belief to improve my work. She is a great listener and always knows how to ask the right questions. I cannot thank her enough for being by my side for the last couple of years not only while I was writing this dissertation but in all aspects of my life.

Contents

List of Figures

CHAPTER 2

CHAPTER 3

CHAPTER 4

Appendices

List of Tables

CHAPTER 1

Introduction

This dissertation is a compilation of publications and publication manuscripts that seek to improve the understanding of measures that are supposed to foster honest behavior both in the private and in the public sector. In these papers, I analyze the effectiveness and mechanisms of prominent measures against dishonest behavior in laboratory experiments. This is of high practical relevance, since the pressure on businesses to promote ethical standards in their organizations has strongly intensified in the last few decades (Dhanarajan, 2005; Waddock et al., 2002). In the course of this development, public officials and regulatory authorities proposed various measures to fight fraud, corruption and other unethical business conduct. Some of the most prominent proposals are the four-eyes principle, monitoring and oath-taking. As a result, these measures are already implemented both in the public and in the private sector (see, for instance, Boatright, 2013; de Bruin, 2016; OECD, 2007; Schickora, 2011). This is potentially problematic because business leaders, stakeholders and public officials rely on the effectiveness of these measures without fully understanding whether and, if so, how they perform in terms of mitigating undesired behavior. In fact, there are several studies that raise doubts about the net positive effects of the cited measures (e.g., Boatright, 2013; Conrads et al., 2013; Fehr & List, 2004; Kocher et al., 2018; Schickora, 2011). Both from a management and from a scientific perspective, this raises several questions: What are effective measures to foster honesty – or, more precisely, which key factors in the decision-making process can be manipulated to effectively increase people's honesty? And what are the mechanisms behind such measures that are indeed able to foster honest behavior?

With this dissertation I aim to contribute to answering these questions. Therefore, I analyze how measures that are supposed to mitigate dishonest behavior perform in a series of laboratory experiments. In these experiments, I focus on lying as one of the most basic forms of dishonest behavior. In line with Mahon (2008), Fallis (2010) and many others, I narrowly define lying as the act of consciously making a false statement.[1] This allows me to examine the basic decision dilemma between the inclination to tell the truth and the temptation to behave dishonestly for personal financial gain. In addition, the focus on lying makes my findings easier to compare to those of other experimental studies on dishonest behavior,

[1] Note that it has been discussed whether lying also requires the intention to deceive another person (e.g., Carson, 1988, 2010; Fallis, 2010; Mahon, 2008). However, the authors of the cited studies agree about the fact that there is always *some* intention behind lying and that this intention depends on the given situation. It can be summarized that a lie is an intentionally made statement that is believed to be untruthful and directed to at least one other person (Bollnow, 1947).

since most of the recent studies in this field of research also focus on lying (e.g., Abeler et al., 2019; Gerlach et al., 2019).

Due to the great variety of measures to prevent dishonest behavior, it would not be practical, if not impossible, to examine all of them at a high level of detail in one single dissertation. To deal with this, my dissertation follows a top-down approach, which consists of the following three steps: The first step is the identification of at least one highly effective measure against dishonest behavior by comparing a broad range of such measures in the lab. The second step is the development of a new experimental design that is suitable for examining the high-potential measure from step 1 in more detail. The third and last step is an analysis of the mechanisms behind the selected measure from step 1 by means of the new experimental design from step 2.

To account for this three-step approach, my dissertations consists of three successively developed and interrelated papers. Each of these papers is self-contained and linked to one of the three steps defined above. Listed below are the three papers with co-authorships and publication status:

- **Can honesty oaths, peer interaction, or monitoring mitigate lying?** (*see Chapter 2*). The first paper is dedicated to analyzing and comparing the effectiveness of distinct measures that are supposed to prevent people from lying for financial gain. This paper has been written in collaboration with Christoph Bühren, Björn Frank and Elina Khachatryan. It has been published in 2020 in the *Journal of Business Ethics*, Volume 163, Issue 3, pages 467-484, https://doi.org/10.1007/s10551-018-4030-z.
- **Size matters! Lying and Mistrust in the Continuous Deception Game** (*see Chapter 3*). The second paper introduces a novel experimental design that allows the measurement of the size of the lie, the size of mistrust, and respective first-order beliefs[2] in a two-player relationship of strategic interaction. It is single-authored and was accepted, after peer review, by the *Verein für Socialpolitik* (VfS) as a conference paper of the VfS Annual Conference 2020. In addition, I submitted the paper to *Theory and Decision* in August 2020.
- **Quick, intuitive truth-telling and deliberate, strategic lying under oath** (*see Chapter 4*). The third paper explores the mechanisms behind the effects of an honesty oath by analyzing its impact on moral deliberation and strategic decision making. This paper is single-authored. It has been accepted for publication by the *Journal of Behavioral and Experimental Economics* in May 2021 (Beck, 2021), https://doi.org/10.1016/j.socec.2021.101728. Note that, at the time of the submission of my dissertation (i.e., in November 2020), the paper was still under peer review.

[2] In game theory, first-order beliefs are expectations one player holds about another player's behavior. On this basis, the first-order beliefs in the two-player game that I introduce in Chapter 3 refer to the size of the lie (or the size of mistrust) that one player expects from the other.

In the following sections, I will elaborate the connection between these three papers and discuss their contribution to the current state of research. Therefore, I will place my work into a broader context and summarize my most important findings. To this end, I will discuss the three papers in their sequential order in the remainder of this introduction (*see sections 1 to 3 of this chapter*). Thereafter, the three papers are presented in full length in Chapters 2 to 4. Each of these papers can easily be read independently. The dissertation concludes with some final remarks, including a short discussion of strengths and limitations of my work as well as suggestions for future research in Chapter 5.

1. Effectiveness of measures against dishonesty

Distinct measures to foster honest behavior

Measures that are supposed to increase people's honesty can be roughly divided into measures that aim at fostering extrinsic and intrinsic motivation to comply with ethical standards (Kreps, 1997). In motivation theory, extrinsic motivation means that "an activity is done in order to attain some separable outcome" rather than for the enjoyment of the activity itself (Ryan & Deci, 2000, p. 60), whereas intrinsic motivation refers to "the doing of an activity for its inherent satisfactions rather than for some separable consequence" (Ryan & Deci, 2000, p. 56). With that in mind, I will now discuss some of the most prominent measures against dishonesty, which largely inspired my work in this dissertation.

One widely used measure that aims at creating an extrinsic incentive to behave honestly is the four-eyes principle. This measure implies that critical decisions always have to be made by groups of at least two persons (UNIDO, 2020). The idea is that the decision makers within the group mutually control each other. While this sounds convincing in theory, Muehlheusser et al. (2015) find no significant difference between the honesty of groups and individuals in a simple lying experiment. Beyond that, it has been argued that groups are less effective in moral decision making due to peer pressure (O'Leary & Pangemanan, 2007) and because guilt and moral costs of lying can be divided between group members (Rose-Ackerman, 1975; Rothenhäusler et al., 2018). In fact, there is broad empirical evidence that groups behave less honestly than individuals (e.g., Conrads et al., 2013; Dannenberg & Khachatryan, 2020; Frank et al., 2015; Schickora, 2011). Despite these findings, the four-eyes principle persists as one of the most frequently used measures to promote ethical decision making (Schickora, 2011).

Another prominent measure against dishonest behavior is monitoring. This measure implies that critical decisions and procedures are supervised by an independent body. What monitoring and the four-eyes principles have in common is that both of them shift the responsibility from the individual to more than one person. The major difference is that the four-eyes principle relies on controlling from within the group, whereas monitoring relies on external controlling. While there is evidence that the presence of another, random person can lead to more truth-telling (Kroher & Wolbring, 2015), less is known about the effects that monitoring has on people's honesty when a monitor is explicitly instructed (and paid) to observe their behavior. On the one hand, monitored decision makers could be motivated to fulfill their duties properly because they might fear punishment or a public decline in trust in case they are caught in dishonest behavior (Kretschmer, 2002). The fact that their behavior is monitored could also direct their attention to moral standards, which could make them avoid being dishonest in order to maintain a positive self-concept (Mazar et al., 2008). On the other hand, monitoring could crowd out the intrinsic motivation to behave honestly by raising the feeling of being mistrusted (Banfield, 1975; Fehr & List, 2004; Kreps, 1997; Schulze & Frank, 2003). Moreover, the monitor and the people they observe could engage in undesired collaboration (Weisel & Shalvi, 2015). Like the four-eyes principle, this measure is widely used in businesses to control unethical behavior.

While these measures provide extrinsic incentives to comply with ethical standards, others aim at fostering intrinsic motivation to behave honestly. One prominent measure is oath-taking. In fact, current management practices have increasingly turned to ethical oaths as an instrument to foster honesty (de Bruin, 2016; Jacquemet, Luchini, Rosaz, et al., 2019). As a result, ethical oaths have been implemented for both entire vocational fields (for instance, the MBA Oath) and entire professions (for example, the Dutch Banker's Oath). Such oaths aim at increasing the decision makers' moral commitment and, therefore, are believed to lead to more honest behavior. However, the effectiveness of oaths as instruments for achieving a desired end has been questioned (Boatright, 2013). Boatright (2013) argues that an "oath is merely a promise or commitment to act in certain ways, but it does not perform any act beyond this" (p. 148). While oaths most certainly cannot fix poorly structured or corrupt systems, they might still give direction for ethical behavior. For instance, Carlsson et al. (2013), Jacquemet et al. (2013), Kemper et al. (2016) and Jacquemet, Luchini, Rosaz, et al. (2019) find that honesty oaths are able to enhance truth-telling in several different settings. In line with this, it has been shown that a formal promise can increase the compliance to pay taxes on time (Koessler et al., 2019). Moreover, Mazar et al. (2008) find that simple moral reminders, such as writing down the Ten Commandments, can lower the tolerance for dishonesty. These findings make oath-taking a promising device to foster honest behavior. But does the honesty oath actually outperform both the four-eyes principle and monitoring?

Experimental comparison of measures against dishonesty

All the above cited studies on the effectiveness of measures against dishonesty focus on one of these measures separately. While this focus on individual measures serves the purpose of these studies in a good way, the results from these studies are sometimes not comparable due to diverging experimental and/or conceptual designs. Therefore, it remains difficult to compare the effects that distinct measures have on people's inclination to tell the truth when given the chance to lie for personal financial gain. In fact, to the best of my knowledge, no previous study compares the effectiveness of the discussed measures in one single experimental design.

Against this backdrop, the paper in Chapter 2 ("Can honesty oaths, peer interaction, or monitoring mitigate lying?") analyses the impact that the discussed measures have on people's lying behavior in the lab. My three co-authors and I published this paper in the *Journal of Business Ethics* in 2020 (Beck et al., 2020).[3] In the paper, we examine the effects of the four-eyes principle, monitoring and an honesty oath on players' lying behavior in the dice experiment of Fischbacher and Föllmi-Heusi (2013).[4] As a complement to the analysis of these measures, we explore how willing people are to lie for others in a reciprocal relationship. Even though our reciprocity conditions are not directly inspired by concrete measures against dishonesty, observing reciprocal lying behavior helps to better understand why people lie in the group conditions, since reciprocity has been shown to affect lying behavior in groups (Barr & Michailidou, 2017). Covering such a broad range of distinct measures allows us to observe a systematic variation of key factors in (dis)honest decision making, such as moral awareness, lie aversion and (shared) moral costs of lying. This enables us to understand which key factors in the decision-making process can be manipulated to effectively mitigate lying. On this basis, we are able to compare the effectiveness of the examined measures and, thus, to identify which one of them has the highest potential to foster honest behavior.

Consistent with studies that found counterproductive effects of the four-eyes principle (e.g., Kocher et al., 2018; Schickora, 2011; Weisel & Shalvi, 2015), our results show that two-player groups lie more than individuals. To better understand the motive to lie here, we are the first to empirically compare the honesty of groups (i.e., when both group members can mutually benefit from lying collaboratively) to the honesty of players who are in a reciprocal relationship (i.e., when lying does not affect the payoff of the liar but the payoff of the other player). By this, we find significantly more lying in the group than in the reciprocity condition, which suggests that group members care more about their own payoffs than about those of their co-players. This implies that group members in our experiment collaborate and lie together mainly for personal financial gain. As a result, we find that the four-eyes principle performs the worst among all examined measures in terms of lie mitigation. However, monitoring barely performs better. In fact, we observe a similar amount of lying in the monitoring condition as in the group condition. Note here that, in contrast to previous studies on monitoring, monitoring in our experiment is not done by necessarily neutral and benevolent authorities. More precisely, the monitors in our experiment are independently paid players who are instructed to report whether another player lies in the dice experiment and to confirm this report with their signature. Since players are unobserved by the experimenters while making their decisions, the monitors are provided with the possibility to collaborate with the players they observe. Under these conditions, we find that most monitors join the lie of their overserved players by covering up untruthful die roll reports. Beyond that, we provide evidence that this

[3] As stated above, my three co-authors for this paper are: Christoph Bühren, Björn Frank and Elina Khachatryan.
[4] In this experiment, participants are given the opportunity to increase their payoffs by lying about the result of a private die roll (Fischbacher & Föllmi-Heusi, 2013). Since each number on a fair six-sided die comes up with the same probability, the true results from multiple die rolls follow a uniform distribution. Hence, lies in this experiment can be detected on a group level by comparing the distribution of reported numbers to a uniform distribution.

form of undesired collaboration happens independently of the size of the monitors' payments. Overall, both the four-eyes principle and monitoring fail to increase people's honesty in our experiment.

These findings challenge the often-assumed net positive effects of controlling measures that aim at creating extrinsic incentives to behave honestly. This is because the expected positive effects of such measures (e.g., an increase in honesty due to social control that arises from the knowledge of being observed, as suggested by Kroher & Wolbring, 2015) can be heavily undermined by undesired collaboration among players or shared guilt within groups (Rothenhäusler et al., 2018).

Finally, we find that, in comparison to the other measures, an honesty oath is most effective in mitigating dishonest behavior in the dice experiment. This finding complements previous studies that show that moral reminders can lead to more honest behavior (e.g., Carlsson et al., 2013; Jacquemet et al., 2013; Jacquemet, Luchini, Rosaz, et al., 2019; Kemper et al., 2016; Mazar et al., 2008). Beyond the scope of these studies, we provide first evidence that an honesty oath can even mitigate partial lies, i.e., lies that do not maximize the liar's payoff. Given that most inexperienced players believe that they can effectively disguise their lies by lying partially (Fischbacher & Föllmi-Heusi, 2013), our findings indicate that one important reason why people do not lie under oath indeed is that the oath fosters their intrinsic motivation to tell the truth.

The comparison of the examined measures against dishonest behavior demonstrates that such measures are more effective if they aim at fostering intrinsic rather than extrinsic motivation to behave honestly. This conclusion is in line with Pruckner and Sausgruber (2013) who show that appealing to honesty is more likely to mitigate dishonest behavior than reminders of legal norms. Moreover, it is consistent with Abeler et al. (2019) who find in a metastudy, which is based on empirical studies from economics, psychology and sociology, that being honest is one of the main motivations for truth-telling.

As we find that an honesty oath outperforms the other examined measures, I analyze the mechanisms behind this measure in more detail in the next steps of my dissertation.

2. Development of a new experimental design to examine the mechanisms behind the effects of oath-taking in strategic interaction

Definition of criteria for the experimental analysis of the mechanisms behind honesty oaths

Given the fact that honesty oaths are frequently used as a management device to foster honest behavior (Blok, 2013; Boatright, 2013; de Bruin, 2016; Jacquemet, Luchini, Rosaz, et al., 2019), it is highly important to better understand not only whether but also *how* such oaths work and which potential side effects they could have on the oath-takers. While there is broad empirical evidence that honesty oaths can lead to more honest behavior (e.g., Chapter 2 of this dissertation and other studies, such as Carlsson et al., 2013; Jacquemet et al., 2013; Kemper et al., 2016), the mechanisms behind them are much less explored. In particular, and even though most management decisions involve strategic interaction, the mechanisms behind oath-taking in a strategic environment have been barely in the focus of research. In fact, the great majority of studies on oath-taking does not involve strategic interaction and the few exceptions do not actually focus on how oath-taking affects it (e.g., Hergueux et al., 2019; Jacquemet, Luchini, Rosaz, et al., 2019). To properly address this issue in an experiment, a suitable experimental design must be found first. For this purpose, I will now define criteria for an experiment that is suitable for analyzing the mechanisms behind honesty oaths when the decision about lying requires strategic thinking: (1) In line with previous studies on honesty oaths, such an experiment should confront the players with the "basic conflict between the temptation to behave dishonestly and the capacity to resist that temptation" (Gerlach et al., 2019, p. 2). This applies to most lying experiments. (2) Moreover, since oaths are usually taken individually, the experiment must allow observation of lying behavior on an individual level. (3) In order to be able to analyze the mechanisms behind oath-taking in a context of strategic interaction, the decision about lying in the experiment must encourage strategic interaction between the potential liar who takes the oath and at least one other person. Hence, the experiment must feature at least two players who are in a strategic relationship. (4) Finally, it has been shown that honesty oaths can have entirely different effects on differently sized lies (Jacquemet et al., 2020). Thus, the experiment should allow for examination of the size of the lie.

To the best of my knowledge, there is no previous experimental design that meets all the above criteria. However, some existing experiments fulfill more of these criteria than others. To show how different experimental paradigms perform in the light of these criteria, I refer to Gerlach et al. (2019) who use similar criteria to compare four of the most widely used experimental paradigms on dishonest behavior: coin-flip tasks, die-roll tasks, matrix tasks and sender-receiver games. According to Gerlach et al. (2019), coin-flip tasks (Bucciol & Piovesan, 2011), die-roll tasks (Fischbacher & Föllmi-Heusi, 2013) and most matrix tasks (Mazar et al., 2008) do not allow examining dishonest behavior on an individual level.[5] Moreover, these three experimental paradigms do not feature strategic interaction. The main reason for this is that lying in such experiments refers to misreporting private information directly to the experimenters (Gerlach et al., 2019). Under these conditions, the liar's expected benefit from lying is clear ex ante and does not depend on another person. Thus, lying in such experiments barely requires strategic thinking. For these reasons, I argue that coin-flip, die-roll and matrix tasks are not suitable for analyzing individual lying behavior in strategic interaction. By contrast, sender-receiver games (Gneezy, 2005) fulfill these two criteria. In such games, lying means sending a false message to another participant

[5] There are some exceptions: Normally, lying in *coin-flip* and *die-roll tasks* is detected on an aggregate level (Gerlach et al., 2019). This is done by comparing the distributions of expected and reported results from privately observed coin flips or die rolls (Bucciol & Piovesan, 2011; Fischbacher & Föllmi-Heusi, 2013). It should be noted, however, that there are computerized versions of these tasks that allow examination of lying behavior on an individual level, since in these versions the coin flip or the die roll can be observed by the computer. As for *matrix tasks*, lying is usually detected by comparing the distribution of reported results from a group that is given the opportunity to cheat to the distribution of results from a group that is not provided with that opportunity (Mazar et al., 2008). However, there also exist versions of this experiment that allow observing individual lying behavior by secretly matching each participant's true result with their reported result (Gerlach et al., 2019).

(Gerlach et al., 2019). Here, the expected payoff of the potential liar (i.e., the sender) depends not only on their own lying behavior but also on the mistrusting behavior of another player (i.e., the receiver). Therefore, sender-receiver games encourage strategic decision making (Sobel, 2009). They also allow observation of lying behavior on an individual level (Gerlach et al., 2019). This makes sender-receiver games more suitable for analyzing individual lying behavior in strategic interaction than the other discussed experimental paradigms.

However, previous sender-receiver games do not allow the measurement of the size of the lie (Gerlach et al., 2019), which would be the last of the criteria defined above. The reason for this is that the senders in such games are confronted with discrete choices.[6] In fact, in most sender-receiver games, the senders have to make the binary choice between lying and truth-telling (e.g., Gneezy, 2005; López-Pérez & Spiegelman, 2013; Peeters et al., 2015; Sánchez-Pagés & Vorsatz, 2007). This becomes a problem when it comes to analyzing the mechanisms behind honesty oaths, since there is evidence that the effects of such oaths on the oath-takers' honesty depend on the size of the lie (Jacquemet et al., 2020).

It can be concluded that sender-receiver games come closest to fulfilling the above criteria for experimental designs to analyze the mechanisms behind oath-taking in strategic interaction – even though they lack the ability to measure the size of the lie. With that in mind, in the paper in Chapter 3 ("Size matters! Lying and Mistrust in the Continuous Deception Game"), I develop a new experimental design that meets all the above criteria. The paper on this experiment was accepted, after peer review, as a conference paper by the *Verein für Socialpolitik* (Beck, 2020) and has recently been submitted to *Theory and Decision*. In the next subsection, I will briefly discuss this new experimental design.

The Continuous Deception Game

The new experimental design, which I develop in Chapter 3, is a sender-receiver game that is framed as an investment game. As it is largely inspired by Erat and Gneezy's (2012) version of Gneezy's (2005) Deception Game, I entitle it *Continuous Deception Game* (CDG). The CDG introduces continuous variables to Erat and Gneezy's (2012) sender-receiver game and, therefore, is the first experimental design that allows the measurement of the size of the lie, the size of mistrust, and respective first-order beliefs on an individual level in a two-player relationship. The sender of this game (i.e., an advisor) is incentivized to lie about the true value of an optimal investment, while the receiver (i.e., an investor) is incentivized to invest optimally. As the receiver does not know the true optimal investment value, they must rely on the alleged optimum reported by the sender. Here, both players have conflicting monetary incentives by design: Whereas the receiver's payoff decreases by any upward or downward deviation of their investment from the true optimal investment, the payoff of the sender increases with the size of the

[6] The only exception that I am aware of is Lundquist et al.'s (2009) sender-receiver game, which allows for different sizes of lies but again features a binary payoff structure for the senders. Therefore, the payoffs that can be expected from two differently sized lies in this experiment are often the same. Hence, the senders' lying behavior can only partially be interpreted on a continuous scale (Gneezy et al., 2018).

receiver's investment. This encourages strategic decision making among both players, since the interplay between their lying and mistrusting behavior has direct financial consequences for both of them. In addition, the continuous message space of both players (i.e., the continuous spectrum of strategies that the players can pursue in the CDG) allows observing more differentiated behavior, which enables testing of more sophisticated theoretical predictions.

Before I use the CDG to examine the mechanisms behind honesty oaths in strategic deception (*see Chapter 4*), I aim to get a better understanding of how both players decide in this new experiment. Having this insight is important not only because I want to use the experiment to examine honesty oaths later, but also because I hope that the CDG will be used to analyze the relationship between (dis)honesty and (mis)trust in other studies that go beyond the scope of my dissertation. Against this backdrop, in the paper in Chapter 3, I also analyze how rational players would behave in the CDG from a game theoretical perspective. Then, I compare these theoretical predictions to the behavior of real players.

On this basis, I find that both players in the CDG tend to make rather strategic decisions. In particular, my results suggest that their behavior is closely related to their beliefs about the other players and that they pursue strategies that are rational from a game theoretical perspective disproportionally more often than other strategies. Furthermore, my results highlight the importance of observing the size of the lie instead of only the proportions of liars and truth-tellers in order to analyze deceptive behavior. This can be illustrated by the finding that the receivers correctly predict the proportion of liars among the senders but largely underestimate the sizes of lies. That is crucial because my results indicate that the senders consciously use the size of the lie to manipulate the receivers. Here, the senders also take advantage of the fact that their first-order beliefs are more accurate than those of the receivers. As a result, I observe that most receivers are manipulated into overinvesting by the senders. These patterns in both players' behavior are largely consistent over time. In addition, the results from my post-experimental questionnaire show that my interpretation of both players' behavior is consistent with their self-assessment of their behavior within the experiment. More precisely, the players consider lying by giving false advice indeed as dishonest, while considering the act of not following received advice as mistrusting. This confirms that the CDG measures what it is supposed to measure.[7]

These findings demonstrate the potential of the CDG as a straightforward method to analyze the relationship between honesty, trust and respective beliefs in strategic deception. Moreover, the

[7] To further improve my understanding of what drives both players' behavior in the CDG, the discussion section of the second paper (*see Chapter 3*) is dedicated to analyzing their decisions in the CDG in the light of the existing literature. By this, I come to the conclusion that the senders mostly lie for financial gain, while altruism and guilt aversion do not seem to play a great role in their decisions. However, I provide evidence that the senders behave more honestly when they expect that their honesty will be rewarded with trust. As for the receivers, my results suggest that their mistrust is largely motivated by risk aversion and expectations of additional monetary gain. In addition, and beyond the scope of previous studies, I find that not only trusting but also mistrusting behavior can be driven by altruism. Finally, I provide evidence of some endogenous preference for trust. For more details, refer to the discussion section of the paper in Chapter 3.

experiment fulfills all the above criteria for examining the mechanisms behind oath-taking in strategic interaction. Against this backdrop, I use the CDG to analyze the mechanisms behind the effects of an honesty oath in the next and last step of my dissertation.

3. The mechanisms behind the effects of an honesty oath

In the paper in Chapter 4 ("Quick, intuitive truth-telling and deliberate, strategic lying under oath"), I use the CDG, which I developed in Chapter 3, to analyze the mechanisms behind an honesty oath similar to the one that has outperformed other measures against dishonesty in Chapter 2. This third paper has been accepted for publication by the *Journal of Behavioral and Experimental Economics* (Beck, 2021). It is inspired by Jacquemet, Luchini, Rosaz, et al. (2019) who show that an honesty oath prior to participation in Erat and Gneezy's (2012) sender-receiver game can change the average decision time of liars and truth-tellers. However, Jacquemet, Luchini, Rosaz, et al. (2019) do not provide empirical evidence of why this happens. With that in mind, the focus of the paper in Chapter 4 lies on causal effects that an honesty oath has on moral deliberation, strategic decision making, and the temporal consistency of (dis)honest behavior. Therefore, I implement a between-subjects design with repeated measurements. Moreover, I examine the decision time to assess the extent of cognitive reasoning and moral deliberation in the decision-making process (Jacquemet, Luchini, Rosaz, et al., 2019; Kahneman, 2011; Rubinstein, 2007; Schunk & Betsch, 2006). In addition, I analyze the impact that the oath has on the relationship between the observed lying behavior and elicited beliefs about others. On this basis, I am able to draw conclusions on the impact of the oath on belief-driven strategic decision making (López-Pérez & Spiegelman, 2013; Lundquist et al., 2009; Peeters et al., 2015). To the best of my knowledge, there are no prior studies that examine this aspect of oath-taking.

Consistent with previous studies (e.g., Beck et al., 2020; Carlsson et al., 2013; Hergueux et al., 2019; Jacquemet et al., 2013), I find that the honesty oath leads to more truth-telling in my experiment. However, in line with Jacquemet et al. (2020), I also find that people who lie under oath barely reduce the sizes of their lies. This indicates that the honesty oath does not even "partially" increase the honesty of liars. For this reason, I will now consider the effects of the oath on lying and on the decision to tell the truth separately.

As for lying, my results show that the oath increases the decision time of liars (which is in support of Jacquemet, Luchini, Rosaz, et al., 2019). Moreover, I provide first evidence that the oath strengthens the relationship between the lying behavior and first-order beliefs of liars. This indicates that the act of lying becomes more strategic after signing an honesty oath, which could explain why liars need more time to decide under oath. These findings reveal a potential negative side effect of oath-taking, namely that the oath could make people who lie under oath deliberate more about how to manipulate others. In some cases, this could lead to better-thought-out lies and, therefore, result in more effective opportunistic behavior. Even though my results strongly imply that the positive effects of oath-taking prevail over such negative side effects, it still seems important to be aware that liars under oath could be inclined to engage in more strategic deception *due* to the oath.

As for the decision to tell the truth, I find that, the quicker people decide under oath, the higher is the chance that they do *not* lie. Beyond that, my results are the first to show that people who are under oath tend to tell the truth regardless of their first-order beliefs. These findings suggest that people decide more intuitively and less strategically under oath. In addition, I provide first evidence that oath-taking makes both the oath-takers' lying and truth-telling behavior more consistent over time.

In the light of the existing literature, these findings improve the understanding of how oath-taking can foster the intrinsic motivation to behave honestly: Referring to Mazar et al. (2008), I argue that taking an honesty oath reminds people of their moral standards by confronting them with the question of what they consider as honest behavior and what not. In most situations, the answer to this question comes to mind without conscious search or computation. In other words, answering this question does not require strategic reasoning but happens largely intuitively (Kahneman, 2003). This implies that truth-telling after taking an oath should be less time-intensive and depend less on strategic beliefs about others, which is exactly what I observed in my experiment. Thus, I argue that an honesty oath encourages the oath-takers to make one basic decision whether to commit to unconditional truth-telling or to deliberately accept being dishonest for personal financial gain. As the outcome of this decision depends on the oath-takers' personal morality, the decision concerns their very identity (Blok, 2013).

Understanding these mechanisms also helps to explain why an honesty oath was more effective in lie mitigation than controlling measures, i.e., the four-eyes principle and monitoring, in the earlier experiment reported on in Chapter 2. In contrast to the honesty oath, such controlling measures do *not* confront people with their own morality (Blok, 2013). Instead, such measures confront people with the question of how others will react to their behavior and which consequences this will have for themselves. Answering this question encourages strategic decision making (Kahneman, 2003) and does not necessarily require moral considerations. Compared to oath-taking, the focus here is less on ethical standards and more on advantages and disadvantages of possible behaviors. This gives people more "wiggle" room to rationalize dishonest behavior. Moreover, by creating extrinsic incentives to comply with rules and ethical norms, controlling measures can undermine the intrinsic motivation to behave honestly (Ryan & Deci, 2000). It can be concluded that, in contrast to an honesty oath, controlling measures lack the ability to make people care about the morality of their actions (Mazar et al., 2008), which can make such measures much less effective in fostering people's honesty than oath-taking.[8]

The following Chapters (*2 to 4*) present the three papers that are part of this dissertation in full text.

[8] A further discussion of this issue can be found in the conclusion of this dissertation (*see Chapter 5*).

CHAPTER 2

Can honesty oaths, peer interaction, or monitoring mitigate lying?[*]

Tobias Beck[♣], Christoph Bühren[♦], Björn Frank[♠], Elina Khachatryan[•]
Institute of Economics, University of Kassel, Nora Platiel-Straße 4, 34127 Kassel, Germany

Abstract. We introduce several new variants of the dice experiment of Fischbacher and Föllmi-Heusi (2013) to investigate measures to reduce lying. Hypotheses on the relative performance of these treatments are derived from a straightforward theoretical model. In line with previous research, we find that groups of two subjects lie at least to the same extent as individuals – even in a novel treatment in which one subject is instructed (and paid independently) to monitor the other. However, we find that our participants hardly lie when lying is only beneficial to others and not to themselves, even if the involved players are in a reciprocal relationship. Thus, we conclude that collaboration on lying mostly happens for personal gain. To mitigate selfish lying, an honesty oath that aims at increasing moral awareness turns out to be effective.

Keywords: Lie detection, Honesty, Moral awareness, Reciprocity, Group decision, Monitoring, Dice experiment.

JEL classification: C91, C92, D63, H26.

Availability of data: The datasets generated and analyzed in this paper are available from the author of this dissertation on reasonable request.

Informed consent: Informed consent was obtained from all individual participants included in the study.

---

[*] A previous version of this paper was circulated under the title “Lying in the face of monitoring, reciprocity, and commitment”, see Abeler et al. (2019) and Garbarino et al. (2017).

[♣] Email: tobi.beck@gmx.net; ORCID: 0000-0002-4162-3648
[♦] Email: buehren@uni-kassel.de
[♠] Email: frank@uni-kassel.de
[•] Email: elina.khachatryan@uni-kassel.de

1. Introduction

Some high-level managers want their employees to lie. A recent example is Wells Fargo where employees were instructed to open millions of unauthorized customer accounts and credit cards (Lynch, 2016). By this, the bank earned unwarranted fees on a large scale. Related to this shady behavior, Wells Fargo fired about 5,300 employees over the last few years (Egan, 2016). Often, however, a firm's management does not want their employees to lie, either because of extrinsic or because of intrinsic motivation. Extrinsic motivation in this case is due to law enforcement, i.e., prosecutors not tolerating a breach of rules against lying in business (see McCord et al., 2004, who emphasize that "lying alone is enough to send you to jail", for a review of cases from the U.S.). Without doubt there is also intrinsic motivation to avoid lying. Prospective managers graduating from the Columbia Business School, for example, have written this in their "Honor Code":

> "As a lifelong member of the Columbia Business School community, I adhere to the principles of *truth*, integrity and respect. I will *not lie*, cheat, steal, or tolerate those who do." (Columbia Business School, 2020, emphasis added).[9]

Honesty oaths like this could hardly persist in business schools if they were completely contrary to students' preferences. But sooner or later after entering a business, weakness of will might give rise to problems regarding honesty and lying, or one needs to keep not only oneself but also others from lying. Thus, from a management perspective the question is: What are good rules to foster honesty? Do honesty oaths work, or does monitoring, or the four-eyes principle?

In this paper, we examine the effectiveness of various measures which are supposed to mitigate lying for financial reasons. By covering a broad range of distinct measures, we observe a systematic variation of key factors in (dis)honest decision making such as moral awareness, moral costs and lie aversion. This allows us to understand the relevance of these factors in the decision-making process, and why measures that aim to raise individual moral awareness (e.g., honesty oaths) so often outperform group-based measures (e.g., the four-eyes principle or monitoring) in other experiments. For that purpose, we report on variations of the dice experiment of Fischbacher and Föllmi-Heusi (2013), in which players have the possibility to cheat in order to influence their payoffs (*see section 2.1.*). Previous studies have shown that signing an honesty oath can increase moral awareness and thus foster honesty (*2.2.*) – we show in our experiment that this works even with partial lies.[10] In addition, we test how the possibility of lying for others influences the players' honesty, which allows us to analyze how players lie for each other in reciprocal relationships (inspired by research briefly reviewed in section *2.3.*). Furthermore, we examine how players lie together in order to test whether the four-eyes principle may or may not work

[9] See also the popular catchword "Hippocratic oath" for business, which is used by many authors, such as Cabrera (2003) and Boatright (2013).

[10] Partial lies are lies that increase the liar's payoff but do not maximize it. For more details, refer to section 2.1.

in terms of lie mitigation. Therefore, we review the literature on group decision making and lying (*2.4.*). Finally, we introduce monitoring by an independently paid observer into the experiment (*2.5.*). By this, we demonstrate that monitoring might not always be an effective measure against dishonest behavior. In sections 3 and 4 we describe our experimental design and procedures, respectively. Section 5 presents a theoretical utility model of costs and benefits of lying, which is motivated by the results of the literature review in section 2. We derive hypotheses from this model and then test them in the results section (*6.*). This is followed by a discussion and a conclusion (*7.*).

2. Literature review

2.1. Lying and the dice experiment

Lies are a strong strategic instrument in human interactions. They are usually used to gain an advantage or avoid a disadvantage in a multiplayer situation. In this paper, we approach lies in a set of experiments based on the dice experiment that was introduced by Fischbacher and Föllmi-Heusi (2013). This experiment is designed to detect and categorize lies in a single-shot decision-making situation. Each participant is instructed to roll a six-sided die in private and report the number they have rolled. Participants are informed that the report will determine their payoff. If the report equals 1, 2, 3, 4 or 5, they receive the corresponding amount in euros (€). If a 6 is reported, the payoff equals €0. Since players are unobserved by the experimenters while rolling their die, they have an opportunity to lie by misreporting. Here, their lies cannot be detected on an individual level. However, since each number on a fair six-sided die comes up with the same probability, the reported numbers could be expected to be approximately equally distributed if all subjects were perfectly honest. Thus, lies can be detected on a group level by comparing the actual distribution of reports to a uniform distribution.

There can be various reasons why a player would choose to lie in this dice experiment. The most obvious reason is the financial benefit that can be gained by reporting an untruthful result (Fischbacher & Föllmi-Heusi, 2013; Gneezy, 2005). For that reason, we use the financial impact of lying to define three categories of lies (Fischbacher & Föllmi-Heusi, 2013; Utikal & Fischbacher, 2013): (1) payoff-maximizing, (2) partial and (3) payoff-reducing lies. Here, a player who is exclusively driven by financial gain always tells a lie of the first category. The second category includes lies that also raise the player's payoff but do not increase it to its fullest possible extent. If a player chooses this course of action, it can be concluded that their decision is driven by at least one non-financial factor. The same applies to players who decide to tell the truth or even lie to reduce their payoff. An ingenious – yet simple – demonstration that various motives play a role for many subjects was provided by Fischbacher and Föllmi-Heusi (2013): They found a disproportionately high number of subjects reporting the number 4, which allows the conclusion that some of them told a partial lie. Moreover, in a study conducted with nuns of a Franciscan community in Germany, some participants intentionally chose to reduce their payoff (Utikal & Fischbacher, 2013) in order to appear honest without being so (Batson et al., 1999). Given the variety of subjects' motives to lie or tell the truth, it is reasonable to hypothesize that variations in institutional design can foster some of these motives and thereby reduce (or increase) lying.

2.2. Moral awareness and honesty oaths

2.2.1. Conceptual ethical work on oaths

Many companies use honesty oaths in form of ethical clauses in order to raise their employees' moral commitment and foster honest behavior by increasing the moral costs of misbehavior (de Bruin, 2016; Jacquemet, Luchini, Rosaz, et al., 2019). This reflects the fact that oaths are a primarily instrumental concept. Historically, they served the purpose of creating a quasi-extrinsic incentive even in the absence of a human norm enforcer: According to Bentham (1817/2005), an oath is "a ceremony composed of words and gestures, by means of which the Almighty is engaged eventually to inflict on the taker of the oath (...) punishment" in the event of breaking the oath (Bentham, 1817/2005, p. 1, cited by Rutgers, 2013, p. 252).[11] Yet according to Boatright (2013), oaths existed even long before religion, since people in primitive times believed in the power of the curse to harm. An oath at that time was simply a *self-curse* that added credibility to the promise of the oath-taker. The workings of moral commitments are harder to explain without religion or primitive believes in magic of the curse. Especially in most contemporary settings, where an external Almighty is no longer involved in an oath, we are dealing with intrinsic rather than extrinsic motivation. Blok (2013) refers to the performativity of an oath: It changes the identity of the swearer,[12] making the breaking of an oath more severe than the breaking of a promise. Arguably an oath is more than a mere promise (Sulmasy, 1999). De Bruin (2016) presents a framework for classifying oaths according to their form, their content and the specific contribution they make to business ethics management. He argues that an oath is made publicly, usually embedded in a ceremonial context, more general than a promise, and commits the oath-taker to behave in a certain way towards particular beneficiaries. Therefore, an oath invokes a moral principal, e.g., honesty, regulations or codes of ethics (van der Linden, 2013). For instance, an ethical oath that is sworn by colleagues can build a common moral ground for collective moral judgment (Hartenberger et al., 2013). According to de Bruin (2016), such an oath "must stipulate norms and values that apply specifically to the oath-taker's situation, or to the profession, if any, for which the oath is designed" (de Bruin, 2016, p. 27). Furthermore, he argues that an oath can have conceptually different but interrelated roles: fostering professionalism, facilitating moral deliberation, or enhancing compliance. We can find these functions of oaths represented in many implementations of oaths in business. For instance, the MBA Oath, which originated from the Harvard Business School in 2009, is a voluntary pledge for MBAs to act ethically

---

[11] This is not to say that the threat of punishment is the only mechanism through which religion works. For example, Lang et al. (2016) show that religious music is a sufficiently effective reminder for religious subjects not to cheat.

[12] Blok (2013) sees ethical oaths as involving "the *intention* of a person not only to *do* something, but also to *be* the one who is committed to some future course of action" (Blok, 2013, p. 193, italics in original).

and responsibly in value creation. Another well-known example is the Dutch Banker's Oath, which was implemented to foster ethical banking practices as a reaction on the financial crisis that began in 2007. Boatright (2013) argues that bankers' oaths do not perform well, since "banking lacks the relevant characteristics of a profession or public office" (Boatright, 2013, p. 161).[13] What all these oaths have in common is that they aim to guide the oath-takers' behavior in a desired direction.

2.2.2. Empirical work on oaths

The papers cited above are conceptual and/or ethical reflections on oaths.[14] However, the effects of oath-taking in concrete settings are unclear ex ante. Hence, there is need for empirical work, which the remainder of this subsection is devoted to.

Mazar et al. (2008) show that directing the attention to moral standards can lower the tolerance for dishonesty. They argue that increasing one's moral awareness might decrease one's self-concept maintenance threshold. Supporting these findings, Pruckner and Sausgruber (2013) show that appealing to honesty is more likely to mitigate dishonest behavior than reminders of legal norms. Moreover, Koessler et al. (2019) find that a promise to pay taxes on time can improve compliance (e.g., payment behavior) in field and laboratory experiments if the promise is linked to a non-financial reward rather than a financial one.

Carlsson et al. (2013), Jacquemet et al. (2013) as well as Kemper et al. (2016) provide evidence that survey respondents asked for their hypothetical willingness-to-pay answer more honestly after taking an oath. In a loaded environment (i.e., when a lie is called a lie), Jacquemet, Luchini, Rosaz, et al. (2019) find that subjects lie less after taking an honesty oath in Erat and Gneezy's (2012) sender-receiver game.[15] Jacquemet, Luchini, Shogren, et al. (2019) show that weak truth-telling in regard to deliberately giving wrong answers in a set of discrete choice experiments happens due to a lack of commitment to tell the truth rather than to a lack of commitment to fulfill the applied duties faithfully and conscientiously. In addition, there is evidence that professional oaths in business may facilitate moral deliberation (de Bruin, 2016). It can be concluded that honesty oaths have been found to reduce lying (at least in the short run) by reinforcing the oath-takers' commitment to the truth in many different experiments. However, to the best of our knowledge, analyses of this measure have not differentiated between its impact on partial and payoff-maximizing lies, which deserves an examination in more detail.

[13] To check the expected impact of the Dutch Banker's Oath, Loonen and Rutgers (2017) conducted a survey among bank employees and their clients. They find that the trust in the Dutch Banker's Oath "does not seem to be very high" (Loonen & Rutgers, 2017, p. 28), and that bank employees even appear opposed to it.

[14] For more extensive reviews of the related literature see, for example, de Bruin (2016), including a comparative evaluation of the MBA Oath, the Economist's Oath (DeMartino, 2011), the Dutch Banker's Oath, and various other similar initiatives, or – with a focus on the private sector, especially banking – Boatright (2013).

[15] In this study, the subjects agreed "to swear upon my honor that, during the whole experiment, I will tell the truth and always provide honest answers" (Jacquemet, Luchini, Rosaz, et al., 2019, p. 429).

2.3. *Lying for others*

In the real world, telling a lie usually also has an impact on other people. With that in mind, the decision maker's reasons for or against lying are probably not the same if lying does not influence exclusively their own but also another player's payoff. In this case, appearing honest may not be as important to the decision maker as being beneficent or helpful. For that reason, a player might lie to increase the other player's payoff. Erat and Gneezy (2012) categorize such behavior as white lies or, in other words, lies that help others. Yet, the decision maker might also lie to reduce the other player's payoff, which is called a black lie or a lie that harms others (Erat & Gneezy, 2012).

Furthermore, Erat and Gneezy (2012) observed "Pareto white lies" (i.e., lies benefiting both the liar and the other player) with a higher probability than lies that are only beneficial to the liar. Indeed, being beneficent has been shown to work as a motivator to drive players to engage in white lies (Gino et al., 2013). However, it must be differentiated whether players tell a white lie because they aim to be (or appear) beneficent or because of plain reciprocity considerations. Barr and Michailidou (2017) show that lying occurs more often if reciprocity is involved by design.

2.4. *Group decision*

Many real-world decisions are also made in groups (Charness & Sutter, 2012; Hoering et al., 2001; Robbins & Judge, 2013). It is commonly agreed that the honesty of groups can differ entirely from that of individuals (see, for instance, Cohen et al., 2009; Danilov et al., 2013). O'Leary and Pangemanan (2007) argue that groups may not be an effective means of producing the optimal decision in ethical decision making due to peer pressure. In particular, groups seem to have stronger self-interested preferences in strategic situations than individuals, even if this may decrease welfare (Charness & Sutter, 2012). This can be explained by the concept of moral costs, which refer to a form of moral doubts and can arise from opportunistic behavior (Rose-Ackerman, 1975; Rothenhäusler et al., 2018). Since these costs are likely to result in guilt, shame or self-reproach, they are generally associated with a negative utility (Abbink, 2000; Kretschmer, 2002; Rose-Ackerman, 1975). As a consequence, one who aims to maintain a positive self-concept will probably try to reduce moral costs (Mazar et al., 2008). In a group, however, these costs can be divided between its members (Rose-Ackerman, 1975; Rothenhäusler et al., 2018). This form of shared guilt makes it easier to maintain a positive self-concept when lying to achieve some kind of mutual financial gain for the group. There is also the effect of conformity with *similar* others (Gino et al., 2009a), hence some people might cheat so as to reinforce the similarity with their group members. In addition, while lifting single players' anonymity makes them less opportunistic, this mechanism is absent for group decisions according to Christens et al. (2019).

Conrads et al. (2013) have already observed these effects in a variation of Fischbacher and Föllmi-Heusi's (2013) dice experiment by comparing the lying behavior of individuals to the lying behavior of two-player groups. In these groups, each player received a die, rolled it in private and reported the result independently of their group member. Afterwards the payoff of each group member was calculated by taking the average of the values that each player would have received in an individual setting. Hence, this form of team incentives made it easier to disguise a lie on an individual level. As a result, groups tended to lie more often than individuals in order to raise their payoff. Similarly, Weisel and Shalvi (2015), Wouda et al. (2017) – whose study is a direct replication of Weisel and Shalvi (2015) with lower effect size – as well as Soraperra et al. (2017) paid two subjects according to their reports of privately rolled dice. The difference to Conrads et al. (2013) is that the second player was informed about the report of the first player before making their decision and that both players were only paid if their reported numbers were identical. This means that the second player in these studies should face a trade-off between honesty and cooperation (collaboration). However, any interpretation of these findings must bear in mind that all players were selected anonymously and had no means to communicate with each other. Thus, no real group interaction was taking place.

Joint decisions normally require some form of interaction (Knebel, 2011; Morgan & Tindale, 2002; Wegge, 2004). At least, the members of the group have to communicate about possible courses of action. This also applies to joint decisions about lying and other forms of dishonest behavior. Kugler et al. (2007) compared the behavior of groups and individuals in a two-person trust game. They found that groups are less trusting than individuals but just as trustworthy. Moreover, Schickora (2011) has examined groups' compliance by conducting a standard corruption game in which players could agree in groups to jointly accept or reject a corrupt transaction. To do so, they were allowed to communicate via an anonymous chat function. The results of Schickora (2011) indicate that group decision making leads to more corrupt behavior than individual decisions. Li et al. (2015) obtain similar results, indicating that individuals exploit moral "wiggle rooms" in groups (Dana et al., 2007). Hence, it can be assumed that in terms of honesty two heads are not necessarily better than one. However, since studies on that subject varied broadly in the extent to which group interaction was permitted, this assumption deserves verification in a context that allows examining the effectiveness of the four-eyes principle when real face-to-face communication is permitted.

2.5. Monitoring

In an economic context that allows opportunistic behavior decisions are likely to be monitored in order to prevent dishonest decisions (Bussmann et al., 2009). However, monitoring can influence people's honesty in various ways. In the first place, it could increase the anticipated opportunistic costs of lying, such as financial punishment or a decline in trust and/or image (Kretschmer, 2002). In addition, the fact of being observed could make the decision makers question the ethics behind their decisions and therefore lead to higher moral costs of lying, such as guilt or shame (Rose-Ackerman, 1975). This in turn could have a negative influence on the self-concept of the potential liar (Mazar et al., 2008). Kroher and Wolbring (2015) observed a similar effect of social control in a dice experiment in which each participant rolled their die in front of another, randomly matched player. The experimental design allowed the other player to observe the die roller's self-reported outcome (they were in a box together). However, the possible observer was *not* explicitly instructed to monitor their co-player. Thus, monitoring happened exclusively out of curiosity. This form of social control was able to reduce the amount of lying, providing evidence that the knowledge of being observed can prevent dishonest behavior.

However, this might change if the observer is explicitly instructed to monitor and report dishonest behavior, since that could trigger a sense of being mistrusted in the observed player, which in turn could lead to the feeling of being suspected of dishonest behavior. As a consequence, the observed player's intrinsic motivation to be honest could be crowded out (Banfield, 1975; Kreps, 1997; Schulze & Frank, 2003). In line with this, Fehr and List (2004) find in two versions of the trust game that incentive structures that are based on the threat of penalizing shirking can reduce trustworthy behavior of students and CEOs in the role of agents. Thus, monitoring backfires in their experiment: The trustworthiness of the agents is the highest if the punishment option is available but not used by players in the role of principals. Nevertheless, most student and CEO principals implement this form of monitoring.

Moreover, the trustworthiness of monitored persons is also likely to depend on which behavior they expect from their observers (Kretschmer, 2002). If an observed decision maker expects that their observer is not inclined to report a detected lie, they do not need to fear punishment or a public decline in trust. This illustrates that the honesty of the monitored decision maker can be influenced by the honesty of the observer. For this reason, it is important that the observer is committed to their duty of reporting undesired behavior truthfully. On the one hand, the observer could be motivated to fulfill their monitoring duty properly in order to maintain a positive self-concept (Mazar et al., 2008). Since they do not have a financial incentive to support dishonest behavior without unofficial arrangements, this could raise their commitment towards their monitoring responsibilities. On the other hand, the observed person could offer the observer a bribe for not reporting dishonest behavior, which would create a possibility of additional financial gain. In addition, the official payment that the observer receives for their monitoring responsibilities could shift their motivation to fulfil their duty from intrinsic to extrinsic and lead to a decline in their commitment (Kretschmer, 2002; Ryan & Deci, 2000; Schulze & Frank, 2003). It can be concluded that the motivation, the moral awareness, and the payment of all players can have a strong impact on the honesty of the decision maker and the observer (Banfield, 1975; Majolo et al., 2006; Mazar et al., 2008; Schulze & Frank, 2003).

The distinction between monitoring and social cues is sometimes vague, as in the impressively simple field experiment of Bateson et al. (2006). They observed payments for tea, coffee and milk made via an "honesty box" in a coffee room shared by forty-eight university staff members. Their treatment variable is a poster placed in the coffee room: In some weeks it showed a pair of eyes, in others flowers. Payments were markedly, and statistically significantly, higher in weeks with eyes, even though there were no sanctions for deviating from the suggested prices.

What all studies in this subsection have in common is the fact that the monitoring is done by benevolent authorities. But what happens if the monitor feels inclined to join in lying? Some of our treatments – described in the next section – allow for this possibility for the first time.

3. Design

In order to analyze the impact of moral awareness, multiplayer interdependences and monitoring on honesty, we modified the dice experiment of Fischbacher and Föllmi-Heusi (2013). Therefore, we compared individual to group behavior with various designs of group interaction: Lying influencing both players' payoffs in the group equally, lying influencing only the co-player's payoff, and lying under monitoring. In addition, we manipulated the moral awareness of individual players by having them sign an honesty oath. As a result, we conducted a total of six treatments:

(1) Baseline treatment
(2) Moral awareness treatment
(3) Direct reciprocity treatment
(4) Indirect reciprocity treatment
(5) Group treatment
(6) Monitoring treatments:
    (a) Unpaid monitoring
    (b) Low-paid monitoring
    (c) Medium-paid monitoring
    (d) High-paid monitoring.

3.1. Baseline

As a control treatment we conducted the dice experiment of Fischbacher and Föllmi-Heusi (2013) in its original form as described in section 2.1. This treatment constitutes a baseline which represents individual behavior and, therefore, is referred to as baseline treatment (or in short: Baseline). At the beginning of this treatment, we explained its procedure to the participants and guaranteed them their anonymity. Therefore, we separated all players from each other and kept a proper distance between them to make clear that they would not be observed during the entire experiment. Then we handed out an instruction sheet to each player which explained the procedure of the treatment. In this sheet each player was instructed to roll one die repeatedly to make sure that it functioned correctly. The players were instructed to memorize and report the first number rolled, as this would determine their payoff in the experiment. The rules for the payoff distribution were explained in a simple payoff table. During the whole experiment we used neutral language and made sure not to mention terms like honesty or lying.

3.2. *Moral awareness*

In another individual player treatment, we aimed at investigating the impact of an increase in moral awareness on players' honesty. In this moral awareness treatment (which is also referred to as: Moral Awareness), all players had to sign an honesty oath in order to raise their commitment to the truth by directing their attention to their moral standards (Jacquemet et al., 2013; Mazar et al., 2008). For the formulation of this oath we used the results from Pruckner and Sausgruber (2013) who find that references to honesty have a stronger impact on people's behavior than references to rules and regulations.[16] What was missing from a traditional oath as described by Bentham (1817/2005) is the element of ceremony, the gestures (*see section 2.2 above*).[17] However, just signing (but not standing up and raising the hand, for example, or repeating after someone else) is just the way some of the present authors have sworn not to take bribes in service of the German federal state of Hesse.

Hence, we emulate common practice when asking subjects in this treatment to confirm with their signature that the data they would provide regarding their actions during the experiment would be in line with the principle of honesty and that they would not lie in order to enrich themselves.[18] Afterwards, we conducted the dice experiment just like in the baseline treatment. Behavioral differences between this treatment and the baseline treatment would allow us to draw conclusions on the short-term effect of an honesty oath on lying while distinguishing between partial and payoff-maximizing lies.

3.3. *Reciprocity*

In two other treatments, we investigated how the possibility of lying for others influences people's honesty in reciprocal relationships. Therefore, we maintained the individual character of the decision-making process but shifted the effect of lying to a multiplayer context. We changed the payment rule so that each player's reported number was not used to determine their own payoff but the payoff of another player. In the direct reciprocity treatment (which is also referred to as: Direct Reciprocity), subjects were informed that their reported number would determine the payoff of another player, and that this player's reported number would determine their own payoff – something that Barr and Michailidou (2017) call "complicity game". In the indirect reciprocity treatment (or in short: Indirect Reciprocity), subjects were informed that their reported number would determine the payoff of another player, and another different player's reported number would determine their own payoff. Deviations of the players' behavior

[16] Nevertheless, Cleek and Leonard (1998) find that a reference to the existence of a corporate code of ethics is as effective as giving details on the code, at least with students imagining to work for a fictitious firm.
[17] There is a good reason for reserving certain ceremonial elements to very special oaths like in court: Rutgers (2010, 2013) warns against the use of honesty oaths in the private sector, as common misuse of those in the private sector might spread into the public sector and harm the meaning that oaths currently hold in the public sector. Different kinds of oaths might reduce the likelihood of such a spillover.
[18] The exact wording of the oath that our participants took in the moral awareness treatment can be found in section 7.6.

between these two treatments would demonstrate differences between direct and indirect reciprocity when potentially engaging in white lies.

3.4. Group decision

Moreover, we conducted a group treatment (which is also referred to as: Group) to compare individual lying to lying based on group decision making. In this treatment, we paired all participants randomly with another player to perform the dice experiment. Since we wanted both players to decide together which number they would report, only one die was rolled per group. The two players had to decide who of them would roll the die.[19] Both were instructed to report the outcome together. Subjects used face-to-face communication in private to come to a joint decision. The amount of each player's payoff was determined by the reported number of the group and was paid out individually. As a result, players of the same group always received the same payoff. In contrast to Conrads et al.'s (2013) and Kocher et al.'s (2018) experiments, the participants in our experiment were not anonymous to their group member. Thus, our setting obviously favored interaction within groups. This allows us to draw conclusions about the impact of real group decision making on honesty in a face-to-face context, and in particular on the effectiveness of the four-eyes principle.

3.5. Monitoring

Beyond that, we conducted a treatment in which one player was instructed to monitor the behavior of another player during the dice experiment. In this monitoring treatment (or in short: Monitoring), we also paired all participants randomly with another player. Here, one of the players was randomly selected to play the dice experiment in front of their co-player. The other player was instructed to observe the numbers that came up when the first player rolled the die and to confirm the result that the first player reported on their group instruction sheet. In order to crowd out possible effects of the observer's payment on both players' honesty (i.e., wage could serve as a reference point for earnings), we varied the observers' payment between subjects. Thus, some observers received no payment (Unpaid monitoring) for their monitoring duty, whereas other observers received €1 (Low-paid monitoring), €3 (Medium-paid monitoring) or €5 (High-paid monitoring). This setting enables us to analyze both the honesty of the observed players and that of the monitors.

4. Experimental procedures

A total of 396 subjects were recruited at the University of Kassel for our experiment – 188 females and 208 males with an average age of 24.25 years. From our post experimental questionnaire, we know that 36% of our participants describe themselves as religious (40% of the females and 33% of the males). 63% of the participants were students of economics and business administration at the University of

[19] Twelve observations were not included in the sample because of a deviation from the experimental protocol. In these cases, the role of rolling the die was not assigned to one group member by the group but by the experimenters.

Kassel, 12% came from educational sciences, 8% from engineering, 7% from various other fields of studies and the remaining 10% were not students at the time of the experiment. We find it an advantage of our subject pool that the majority of the students in our sample have an economics background, as such students are most likely to be the future newcomers in areas of the private sectors in which business ethics might be most relevant, such as banking or finance (Teixeira & Rocha, 2010). Due to the brevity and simplicity of our game, we were able to recruit subjects for our "add-on" experiment at the end of several other unrelated experiments[20] as well as a graduation ceremony. Our data collection phase took place between July 2014 and December 2016.

The average payoff for all of our participants was €3.16, which is well above the average hourly rate for a student, since the experiment took only a few minutes to complete.

5. Theoretical model of the utility of lying

In a basic model, utility U_i of individual i depends on

i's monetary payoff p_i,
and others' monetary payoff p_j (thus allowing for social preferences)

with $dU_i / dp_i > 0$, $dU_i / dp_j \geq 0$, and $i \neq j$. Because of diminishing marginal returns of money, we might have:

$$U_i = \sqrt{p_i} + m\beta\sqrt{p_j} \tag{1}$$

with $0 \leq ß \leq 1$ indicating how much individual i cares about others and m denoting the number of other persons who get a payoff. If $ß = 0$, i's utility depends solely on their own monetary payoff and not on the payoffs of others: In this case, i would be a *homo oeconomicus*. While we believe that $ß > 0$, we also assume that $\beta \leq 1$ because we argue that it is unlikely that subjects care more for others than for themselves (in analogy, e.g., to the model of Fehr & Schmidt, 1999).

When we introduce the possibility of lying for oneself (L_i) or for another person j (L_j), the payoffs in the dice experiment are calculated by $p_i = r_i + L_i$ and $p_j = r_j + L_j$, where r_i denotes the actual die roll for oneself and r_j denotes the actual die roll for another person. However, lying might induce moral costs C_i. The moral costs of lying for oneself (i.e., to increase the own payoff) can be different from the moral costs of lying for someone else (i.e., to increase another person's payoff). Thus, we have $C_i = \delta_i L_i + m\delta_j L_j$ with $\delta_{i,j} \geq 0$ ($\delta_{i,j}$ would be zero for individuals who absolutely do not mind lying).

[20] Amongst other dice experiments, that of Fischbacher and Föllmi-Heusi (2013) has also been conducted that way.

Usually, we will have $\delta_i > \delta_j$ because the costs of lying for someone else are diluted by the norm of helping others.[21]

If the responsibility for the decision can be shared, the moral costs of lying are divided by the number n of people present when making the decision (e.g., $n = 2$ in our group and monitoring treatments). Furthermore, the moral costs of lying can be made more explicit (or salient) by the awareness factor $\alpha \geq 1$, e.g., by an honesty oath (in the moral awareness treatment) or by an independent observer (in the monitoring treatments). If a treatment does not address the salience of the moral costs, we assume $\alpha = 1$ for simplicity.

Taken together, the utility U_i with the possibility to increase payoffs by lying evolves to

$$U_i = \sqrt{r_i + L_i} + m\beta\sqrt{r_j + L_j} - \frac{\alpha(\delta_i L_i + m\delta_j L_j)}{n}. \tag{2}$$

Maximizing (2) with respect to L_i, we get:

$$L_i = \left(\frac{n}{2\alpha\delta_i}\right)^2 - r_i\,. \tag{3}$$

Maximizing (2) with respect to L_j, we get:

$$L_j = \left(\frac{n\beta}{2\alpha\delta_j}\right)^2 - r_j\,. \tag{4}$$

In our group treatment, we have $r_i = r_j$ and $L_i = L_j$. Maximizing (2) with respect to L_i in the group treatment leads to:

$$L_i = \left(\frac{n(1 + m\beta)}{2\alpha(\delta_i + m\delta_j)}\right)^2 - r_i\,. \tag{5}$$

These three equations have intuitive interpretations: The optimal amount of lying for someone else increases with β – in words: If one cares about others' wellbeing and lying can help others, one lies more. Moreover, the optimal amount of lying increases with n, i.e., with the number of people who share the moral costs of the lie. Furthermore, the stronger the dislike of lying for oneself (δ_i) or for another person (δ_j), and the higher the salience of the costs of lying α (e.g., by swearing beforehand not to lie), the less one lies. Finally, individual i has less "space" for lying if their die roll (that is supposed to determine their individual and/or another person's payoff) yields a higher result.

[21] Wu et al. (2011) provide neuro-economic evidence that engagement in dishonesty purely for the benefit of others can be perceived as morally acceptable. This study supports our assumption that δ_j does not just consist of moral costs of cheating (equal to δ_i) but is indeed diluted by the norm of helping others.

A straightforward test would require a continuous variation of parameters of the utility function. While this would probably be impossible to implement in an experiment, we have treatments in which certain parameters or variables of the utility function U_i are zero by design, as shown in Table 2-1.

Treatment	Baseline	Moral Awareness	Direct Reciprocity	Indirect Reciprocity	Group	Monitoring
# of others getting paid by the decision	$m = 0$	$m = 0$	$m = 1$	$m = 1$	$m = 1$	$m = 0$
# of subjects watching the die roll	$n = 1$	$n = 1$	$n = 1$	$n = 1$	$n = 2$	$n = 2$
Own payoff	$p_i \geq 0$	$p_i \geq 0$	$p_i = 0$	$p_i = 0$	$p_i \geq 0$	$p_i \geq 0$
Other's payoff	$p_j = 0$	$p_j = 0$	$p_j \geq 0$	$p_j \geq 0$	$p_j \geq 0$	$p_j = 0$
Moral costs	$\delta_i L_i$	$\alpha\delta_i L_i$	$\delta_j L_j$	$\delta_j L_j$	$\frac{(\delta_i + \delta_j)L_i}{n}$	$\frac{\alpha\delta_i L_i}{n}$
Optimal amount of lying	$\left(\frac{1}{2\delta_i}\right)^2 - r_i$	$\left(\frac{1}{2\alpha\delta_i}\right)^2 - r_i$	$\left(\frac{\beta}{2\delta_j}\right)^2 - r_j$	$\left(\frac{\beta}{2\delta_j}\right)^2 - r_j$	$\left(\frac{1+\beta}{\delta_i + \delta_j}\right)^2 - r_i$	$\left(\frac{1}{\alpha\delta_i}\right)^2 - r_i$

Table 2-1. Parameters affecting lying behavior by treatment

Based on our model, we can develop the following *hypotheses*:

Compared to Baseline, we expect the amount of lying to be…

...lower in Moral Awareness, since the awareness parameter $\alpha = 1$ in Baseline and $\alpha \geq 1$ in Moral Awareness (***Hypothesis H1***).

…lower in Direct and Indirect Reciprocity (***Hypothesis H2a***). Comparing the optimal amount of lying in the reciprocity treatments to Baseline, we have two opposing effects: On the one hand, we assume that individuals tend to care more for themselves than for others ($\beta \leq 1$), which could cause them to lie less in the reciprocity treatments than in Baseline. On the other hand, the moral costs of lying for others are assumed to be lower than the costs of lying for oneself ($\delta_i > \delta_j$), which could cause the players to lie more in the reciprocity treatments than in Baseline. However, we expect the former effect to be larger than the latter. Beyond that, we expect the amount of lying to be smaller in Indirect Reciprocity compared to Direct Reciprocity (***Hypothesis H2b***) because the moral costs parameter (δ_j) should be lower in Direct Reciprocity than in Indirect Reciprocity, since the norm of helping others is more salient in Direct Reciprocity.

…higher in Group (***Hypothesis H3***) because $1 + ß > 1$, $n = 2$, and $\delta_i + \delta_j < 2\delta_i$, which means that lying is good for oneself and for the other group member, the moral costs of lying are divided by 2, and lying for someone else is less costly than for oneself.

…higher in Monitoring (***Hypothesis H4***), since the moral costs of lying are divided by 2 in Monitoring (*n = 2*). This should have a greater impact on the amount of lying than the increase in the die roller's awareness of the moral costs of lying due to monitoring when compared to Baseline ($\alpha \geq 1$). Moreover, we expect that α is lower in Monitoring than in Moral Awareness because α is affected extrinsically in Monitoring and intrinsically in Moral Awareness (*see section 2.5*). Against this backdrop, we expect that $\alpha < 2$ in Monitoring.

In addition, the amount of lying should be lower in Moral Awareness compared to Monitoring (***Hypothesis H5***), since even if α is assumed to be equal in both treatments, the moral costs of lying are divided by 2 only in Monitoring ($n = 2$).

Moreover, we expect to observe more lying in Group compared to Monitoring (***Hypothesis H6***) because players in Group can mutually benefit from lying (with $ß > 0$) and because we assume that being observed by and independent monitor could make the moral costs of lying more salient in Monitoring ($\alpha \geq 1$).

Finally, we predict a higher amount of lying in Group compared to the reciprocity treatments (***Hypothesis H7***) because lying in Group can not only increase the other player's payoff but also the payoff of the liar ($1 + ß > ß$). While the overall moral costs of lying are generally higher in Group than in the reciprocity treatments ($\delta_i + \delta_j > \delta_j$), they are divided by 2 only in Group ($n = 2$).

6. Results

Table 2-2 summarizes the average payoffs by treatment. Without lying the expected average payoff in our experiment would be €2.50. However, the average payoff was significantly higher than €2.50 in all treatments, except for Moral Awareness and Indirect Reciprocity (according to separate one-sample Wilcoxon signed rank tests). In these two treatments, subjects earned on average significantly less than in Baseline (separate two-sided Mann-Whitney U tests: $p < 0.05$). This suggests that swearing not to cheat and rather complex reciprocity relations[22] were able to reduce (or even eliminate) lying. Overall, lying differed significantly between treatments (Kruskal-Wallis test: $p < 0.001$). In the pooled monitoring treatments, the average payoff was only slightly higher than in Baseline, indicating a similar amount of lying in both of these settings.

[22] Note that neither a Kolmogorov-Smirnov test ($p = 0.317$) nor a two-sided Mann-Whitney U test ($p = 0.178$) finds significant differences between Direct Reciprocity and Indirect Reciprocity.

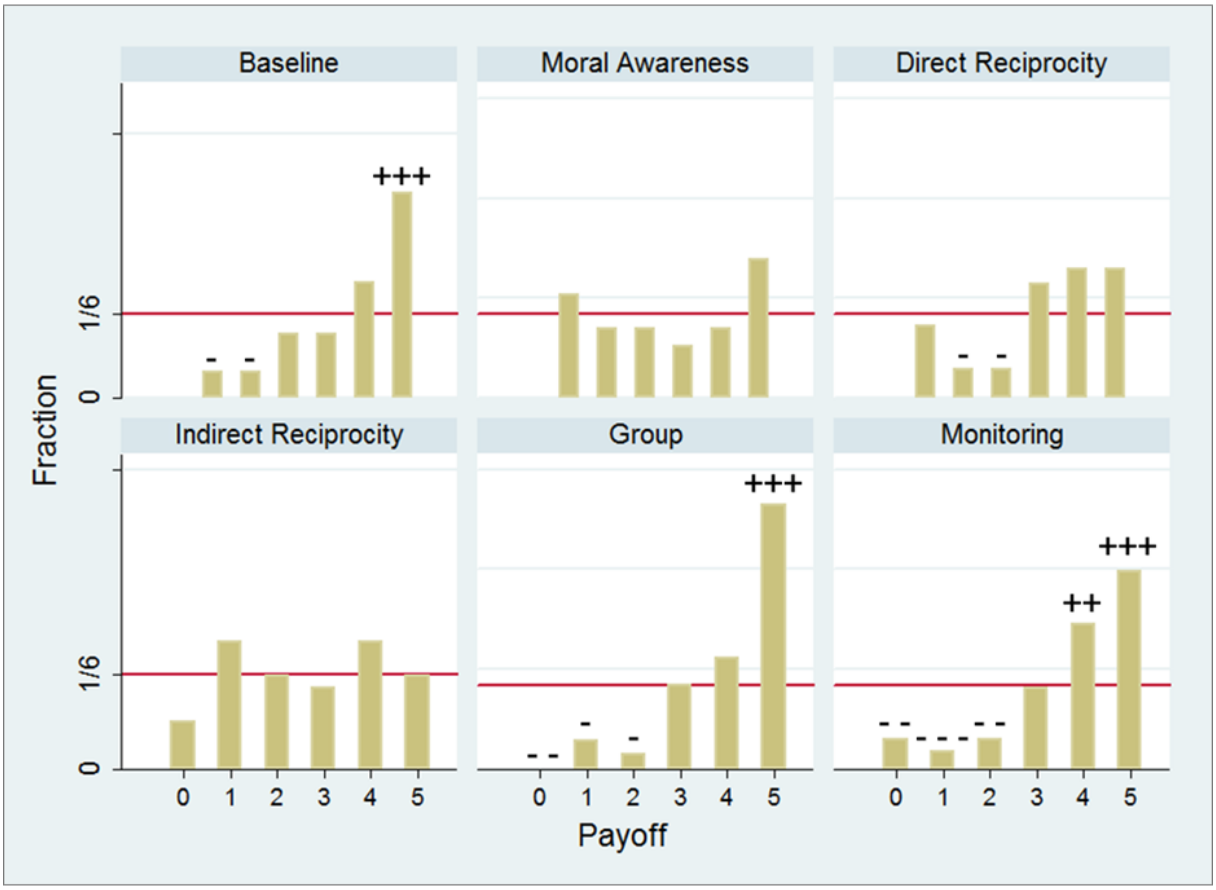
Treatment	Number of observations	Number of subjects	Average payoff (€)	Std. dev. (in €)	Proportion of €5 (in %)
Baseline	39	39	3.67	1.51	41.03
Moral Awareness	29	29	2.66*	1.95	27.59
Direct Reciprocity	35	35	3.17	1.71	25.71
Indirect Reciprocity	49	49	2.69**	1.62	16.33#
Group	36	72	4.14	1.15	52.78
Monitoring	86	172	3.78	1.42	39.53
Unpaid Monitoring	*15*	*30*	*3.47*	*1.77*	*40.00*
Low-Paid Monitoring	*39*	*78*	*3.82*	*1.27*	*35.90*
Medium-Paid Monitoring	*16*	*32*	*3.81*	*1.83*	*56.25*
High-Paid Monitoring	*16*	*32*	*3.94*	*0.93*	*31.25*
Total	274	396	3.42	1.62	34.31

Note. Kruskal-Wallis test that analyzes whether the mean ranks of payoffs are equal across treatments: $p < 0.001$. Two-sided Mann-Whitney U test to compare the mean rank of payoffs in a specific treatment to that in Baseline: *: $p < 0.05$, **: $p < 0.01$. Two-sided Fisher exact test that tests whether the proportion of the highest payoff (€5) differs from that in Baseline: #: $p < 0.05$. Due to the fact that the dice experiment is usually an "add-on" experiment, we have an uneven number of observations between treatments. Differences between number of subjects and observations: In the monitoring treatments, we have 86 players who get paid for monitoring but not for rolling the dice (i.e., the monitors are not included in the number of observations). In the group treatment, in each group two subjects agree on the number they report (i.e., the number of observations is half of the number of subjects).

Table 2-2. Average payoff by treatment

Figure 2-1 shows the distributions of payoffs by treatment. As can be read from the figure, we paid the highest amount (€5) significantly too often in Baseline, Group, and Monitoring.[23] In the reciprocity treatments and in Moral Awareness, high payoffs were not significantly overrepresented (separate one-sided binomial tests are not significant). Yet, for every treatment, we can reject the null hypothesis that the payoffs follow a uniform distribution, which is the distribution we would expect from rolling dice without lying about the outcome (separate one-sample Kolmogorov-Smirnov tests: $p < 0.05$, except for Indirect Reciprocity, where the distribution is closest to a uniform distribution with $p = 0.063$).

[23] In the pooled monitoring treatments, the two highest payoffs were significantly overrepresented.



Note. One-sided binomial tests that test whether the observed fraction of players who received the respective payoff is smaller resp. higher than 1/6: -/+: p < 0.05, --/++: p < 0.01, ---/+++: p < 0.001.

Figure 2-1. Payoff distribution by treatment

In the following paragraphs, we describe the results of our treatments in more detail.

The results of our *baseline treatment* are similar to those of Fischbacher and Föllmi-Heusi (2013). Obviously, the payoffs in this treatment are not equally distributed (one-sample Kolmogorov-Smirnov test: $p < 0.001$). To be more specific, the payoff distribution suggests that some players cheated in order to raise their payoff, i.e., some players who would have received relatively low payoffs engaged in payoff-maximizing lies (according to separate one-sided binomial tests reported in Figure 2-1).[24]

In our *moral awareness treatment*, the average payoff is close to the average payoff of an equal distribution. Yet, the payoffs in this treatment are not equally distributed (one-sample Kolmogorov-Smirnov test: $p = 0.024$), which indicates that some players who signed the honesty oath might still have cheated. However, subjects in this treatment received significantly lower payoffs than in Baseline (€2.66 vs. €3.67, two-sided Mann-Whitney U test: $p = 0.035$), supporting our first *hypothesis H1*. That means that honesty oaths were able to significantly reduce payoff-increasing lies. As a result, we could

[24] In our post-experimental questionnaire, we asked participants, inter alia, for their age. For Baseline, we find a significant negative correlation between players' payoff and age (Spearman's rank correlation: $\rho = -0.512$ with $p = 0.002$). Hence, players' payoffs decrease with increasing age.

neither detect partial nor payoff-maximizing lies in this treatment (separate one-sided binomial tests are not significant).

In our *direct reciprocity treatment*, the payoffs are also not equally distributed (one-sample Kolmogorov-Smirnov test: $p < 0.001$). The €1 and €2 payoffs are significantly underrepresented (separate one-sided binomial tests: $p < 0.05$). Moreover, we do not find significant differences between this treatment and Baseline (two-sample Kolmogorov-Smirnov test: $p = 0.780$; two-sided Mann-Whitney U test: $p = 0.177$), which speaks against *hypothesis H2a*.

This changes in our *indirect reciprocity treatment*. Subjects earned significantly less in this treatment than in Baseline (€2.69 vs. €3.67, two-sided Mann-Whitney U test: $p = 0.004$), which is consistent with *hypothesis H2a*. This shows that players are less willing to lie for another person than for themselves. Compared to Baseline, we observed a significantly lower proportion of the highest payoff (€5) in Indirect Reciprocity (41.03% vs. 16.33%, two-sided Fisher exact test: $p = 0.015$) with the percentage of the highest payoff in Indirect Reciprocity being close to the expected value of 16.67% without lying. In addition, the average payoff in this treatment is close to the expected payoff under complete honesty, which suggests that the separation between decision maker and beneficiary without direct reciprocity significantly reduced payoff-increasing lies. Comparing the mean rank of payoffs between Indirect and Direct Reciprocity, we do not find a significant difference (two-sided Mann-Whitney U test: $p = 0.178$), i.e., *hypothesis H2b* is not (significantly) supported by our data. Pooling the *reciprocity treatments*, the amount of lying in these treatments is significantly lower compared to Baseline (*Hypothesis H2a*, €2.89 vs. €3.67, $p = 0.011$) and compared to Group (*Hypothesis H7*, €2.89 vs. €4.14, $p < 0.001$).

In our *group treatment*, the lying behavior was similar to that in Baseline. Like in Baseline, we detected payoff-maximizing lies (one-sided binomial test: $p < 0.001$). Moreover, we do not observe a significant difference between the payoff distribution in this treatment and in Baseline (two-sample Kolmogorov-Smirnov test: $p = 0.810$). These findings indicate that the four-eyes principle was not able to mitigate lying. On the contrary, the average payoff was even higher for groups than for individuals, which is consistent with *hypothesis H3*. However, this difference is not significant (€4.14 vs. €3.67, two-sided Mann-Whitney U test: $p = 0.189$).[25]

The same can be said for our *monitoring treatments*. Here, some players lied partially or even payoff-maximizing (separate one-sided binomial tests reported in Figure 2-1). Again, we cannot detect significant differences between the pooled monitoring treatments and Baseline (two-sample Kolmogorov-Smirnov test: $p = 0.996$; two-sided Mann-Whitney U test: $p = 0.793$). Thus, *hypothesis H4*

[25] Moreover, the dishonesty of groups appears to be connected to the degree of acquaintance between their group members: In a pre-test for this treatment, we asked subjects to rate the acquaintance with their co-player on a 7-point-scale between “unknown” and “very close”. We found that the degree of acquaintance is highly correlated with the groups’ payoff (Spearman’s rank correlation: $\rho = 0.644$ with $p < 0.001$). In addition, “very close” groups earned on average significantly more than other groups (two-sided Mann-Whitney U test: $p = 0.002$). This suggests that groups behave less honestly, the more familiar their group members are.

is not supported by our results. The average payoffs with and without monitoring are barely different, which suggests that monitoring did not reduce the amount of cheating. Moreover, lying in the monitoring treatments did not change with the payment of the monitor (Kruskal-Wallis test: p = 0.787; Jonckheere-Terpstra test: p = 0.583). Thus, the monitor's payment does not seem to have served as a reference point. In addition, most of the monitors were willing to cover up the lies of their die rollers: 84 out of 86 monitors reported that their co-player had told the truth. Taking into account only the 84 pairs of players where the monitor did not report a lie, the payoffs in the pooled monitoring treatments still do not follow a uniform distribution (one-sample Kolmogorov-Smirnov test: p < 0.001). This indicates that some of the monitors joined the lie of their co-player. Comparing the pooled monitoring treatments and Moral Awareness, we find that the average amount of lying is significantly lower in Moral Awareness, supporting *hypothesis H5* (two-sided Mann-Whitney U test: p = 0.009). Yet comparing the average payoffs in Monitoring and Group, we do not find a significant difference (p = 0.186), i.e., *hypothesis H6* is not (significantly) supported by our results. Table 2-3 summarizes the results of our hypotheses testing.

Hypothesis	Prediction about the amount of lying by treatment	Difference (in €) of average payoff (first - second treatment)	*p*-value (two-sided Mann-Whitney U test)	Result
H1	Baseline > Moral Awareness	1.01	0.035	Supported
H2a	Baseline > Reciprocity (Direct + Indirect)	0.78	0.011	Supported
H2b	Direct Reciprocity > Indirect Reciprocity	0.48	0.178	Not supported
H3	Group > Baseline	0.47	0.189	Not supported
H4	Monitoring > Baseline	0.11	0.793	Not supported
H5	Monitoring > Moral Awareness	1.12	0.009	Supported
H6	Group > Monitoring	0.36	0.186	Not supported
H7	Group > Reciprocity (Direct + Indirect)	1.25	< 0.001	Supported

Table 2-3. Summary of hypotheses testing

Since our theoretical *model of the utility of lying* was able to predict some major differences between our treatments correctly, we use this model to interpret our findings in more detail. For that reason, we estimate the most essential factors of the model using the observed lying behavior (*see appendix A.*). This allows us to draw some final conclusions: (1) The honesty oath increased the moral awareness of our participants by about 20% on average. (2) The dislike of lying for oneself was at least 15 times higher than the dislike of lying for another person. (3) This was due to the fact that the dislike of lying for another person was weak. (4) Moreover, the degree to which players cared about others in the experiment was low. And, finally: (5) Monitoring increased the dislike of lying for oneself by at least 70%.[26] However, potential positive effects of monitoring were still overshadowed by the division of moral costs between both players.

[26] This approximation already captures the potential effects of an increase in players' moral awareness on their honesty due to monitoring.

Lastly, it should be noted that in our post-experimental questionnaire we asked subjects, inter alia, whether they are religious or not. While we find no significant differences in religiosity by treatment, our results show that on average individuals and groups[27] earned less if they stated that they are religious (€3.13 vs. €3.61, two-sided Mann-Whitney U test: p = 0.035). However, in Direct Reciprocity, players with a non-religious co-player received significantly lower payoffs than players with a religious co-player (€2.73 vs. €4.33, two-sided Mann-Whitney U test: p = 0.011). In fact, we cannot detect any cheating by non-religious subjects in this treatment (neither a one-sample Kolmogorov-Smirnov test nor separate one-sided binomial tests for all payoff values are significant). By contrast, subjects with a religious co-player received the highest payoff of €5 significantly too often (one-sided binomial test: p = 0.009). These findings indicate that religious players lied in order to maximize their direct co-player's payoff, whereas non-religious subjects reported truthfully in the same context. In the light of this, it could be speculated that the moral costs of lying for someone else (δ_j) were lower among religious subjects, possibly because religious beliefs so strongly encourage the social norm of helping others. Beyond that, in Indirect Reciprocity, we observed a similar but not significant effect. Here, subjects with a non-religious co-player earned moderately but not significantly less than players with a religious co-player (€2.52 vs. €3.22, two-sided Mann-Whitney U test: p = 0.134).

7. Discussion and Conclusion

This work was originally inspired by the finding that group decision making increases rather than decreases corruption in the experiments of Schickora (2011) and Li et al. (2015). It is a question of utmost practical relevance for business as well as for politics whether there are any conditions that can make the four-eyes principle perform better than in these two experiments. We turned to the dice experiment of Fischbacher and Föllmi-Heusi (2013), on the one hand to check the robustness of the mentioned corruption experiment results, and on the other hand as a means of examining honesty in different variants of group design. However, the dishonesty of groups appears to be a persistent phenomenon. In order to answer the question of which factors are essential for (dis)honest decision making, we introduced our theoretical model of the utility of lying. In this model, we included a broad range of distinct measures, such as moral awareness, the aversion of lying for oneself and/or for others' well-being, the degree to which players care for others, and the division of moral costs in group decision making. Using this model, we systematically varied these key factors in (dis)honest decision making throughout our experiments. This allows us to conclude in more detail which factors are most relevant when one aims to mitigate cheating.

7.1. Group decision

To begin with, our group decision results indicate that groups with real face-to-face communication do not behave more honestly than individuals. In this respect, our results are in line with those of Kocher

[27] We defined religious groups as groups in which both subjects state that they are religious.

et al. (2018) who argue that groups lie more than individuals because they are exposed to arguments in favor of norm violation when discussing their decision. Beyond that, we use our theoretical model of the utility of lying to further analyze the reasons why groups lied in our experiment. On this basis, we estimate that the players' dislike of lying for their co-player was weak. As a result, lying for another person *while* lying for oneself barely raised the moral costs of lying. Moreover, we estimate that the players did not care much about their co-player's payoff. This leads us to the conclusion that cheating in groups was hardly enhanced by altruistic reasons as defined by Erat and Gneezy (2012). In addition, our theoretical utility model allows us to estimate that the division of moral costs between group members was the main reason for dishonest behavior in groups. Our findings support those of Weisel and Shalvi (2015) who argue that potential liars face a trade-off between collaboration and honesty (*see section 2.4 for a detailed discussion*). This demonstrates how real face-to-face communication can result in dishonest decision making by favoring unfaithful cooperation.

7.2. Monitoring

Similar (but yet slightly different) mechanisms seem to be at work when the group is divided into a monitor and an observed decision maker (*see section 2.5*). Regarding this issue, Gino et al. (2009b) implement a monitor in the matrix task of Mazar et al. (2008). When communication was permitted in Gino et al.'s (2009b) experiment, the probability of lying decreased. The probability was much higher, however, when communication between the monitor and the monitored person was encouraged. Our results confirm that the amount of lying is not reduced by monitoring when face-to-face communication is allowed. Furthermore, we use our theoretical utility model to approach the answer to the question of why monitoring so often fails to reduce dishonest behavior. Therefore, we estimate the impact of monitoring on the dislike of lying for oneself. According to our estimates, monitoring increased this form of lie aversion in our experiment by at least 70%. We interpret this number as the impact of social control on honesty that can arise from the knowledge of being observed (Kroher & Wolbring, 2015). In contrast to Kroher and Wolbring (2015), however, the observing players in our experiment were explicitly instructed to monitor their co-players. In this context, our results illustrate that potential positive effects of social control (as observed by Kroher & Wolbring, 2015) can be overshadowed by counteracting effects caused, for example, by the possibility of dividing moral costs of lying between the monitor and the observed player. It is noteworthy that we find that most monitors joined the lie of their co-player by covering up his or her cheating. This gives a deep insight into how strongly undesired collaboration can negatively affect honesty.

The results of our group and monitoring treatments may warn us against the unethical collusive potential of teams and/or principal-agent pairs. Especially in the C-suite, this could have serious implications when, for instance, CEO and CFO or Chairman and CEO collaborate in unethical behavior.[28] Recent

[28] We owe this point to an anonymous referee.

scandals, inter alia, the diesel deception of Volkswagen (see, for instance, Jung & Park, 2017), indicate that groups of powerful decision makers collaborated in cheating and lying.

Yet all is not lost.

7.3. Moral awareness

Our moral awareness results show that an honesty oath can mitigate lying when signed immediately before the moral decision in question. This is consistent with the results of Mazar et al. (2008), Carlsson et al. (2013), Jacquemet et al. (2013), Kemper et al. (2016), and Jacquemet, Luchini, Shogren, et al. (2019) who show that moral reminders can lead to more honest behavior (*see section 2.2*). Beyond the scope of these studies, we find that an honesty oath can even mitigate partial lying. This means that the oath in our experiment was able to prevent players from lying in disguise as introduced in the dice experiment by Fischbacher and Föllmi-Heusi (2013). Moreover, by using our theoretical utility model, we estimate the short-term effect that the oath had on moral awareness in our experiment. On this basis, we conclude that moral awareness increased by about 20% due to the oath. This is interesting because it indicates that the increase of moral awareness that is necessary to prevent people from telling (even disguised) lies is not that high. This may explain why simple reminders of moral standards are so effective in fostering honest behavior in other empirical studies (such as, for instance, Mazar et al., 2008).

However, since moral awareness is expected to weaken over time, the long-term effects of honesty oaths are much more debatable.[29] Another concern that is beyond the scope of our experiment is the multifacetedness of many agents' tasks, which is difficult to reflect in an oath (Boatright, 2013). For example, the "Financial Hippocratic Oath" lets bankers swear to "seek fair fees and appropriate compensation"[30], which opens the door to self-serving biases. This concern is in line with Shalvi et al. (2015) who argue that ambiguity is often exploited in order to justify immoral behavior. An oath that focuses on one dimension of ethical behavior when various dimensions are relevant might also serve as a moral license for wrongdoings not explicitly referred to in the oath (with respect to CSR policies instead of honesty oaths, this has been observed in a field experiment of List & Momeni, 2020). In economic experiments on oaths and lying (including our own), we have a clear understanding of what truthful behavior is, or what exactly constitutes the breaking of an oath. This divergence between practical and experimental oaths indicates that honesty oaths outside of the lab face the challenge of covering a broad spectrum of immoral behaviors in a multitude of possible situations on the one hand while explicitly addressing what moral behavior is in each of these situations on the other hand.

[29] On that subject, Dan Ariely (interviewed by Haas, 2016) states that players hardly cheat when they are reminded of their moral values, but that they do not remember them even the very next day.

[30] Quoted from Boatright (2013, p. 151), who also provides a careful discussion of the less straightforward problems in real bankers' oaths.

7.4. Reciprocity

Honesty seems to depend on the impact lying has on others (*see section 2.3*). In this respect, Gino et al. (2013) provide evidence that individuals lie more when others can benefit from their lying behavior in the matrix task of Mazar et al. (2008). However, Gino et al. (2013) observed significantly less cheating when lying was only beneficial to others – a finding supported by the results of our reciprocity treatments. To this we add that the willingness to engage in white lies is seemingly weaker when direct reciprocity considerations are ruled out. Moreover, we contrast the dislike of lying for oneself with the dislike of lying for others. To estimate these two forms of lie aversion, we compare the lying behavior in our reciprocity treatments to the lying behavior both in our baseline and in our group treatment. Based on these comparisons and by using our theoretical utility model, we estimate that the dislike of lying for others in our experiment was weak. However, this did not result in a high amount of lying for others when the players were in a reciprocal relationship. According to our utility model, the reason for this is that our subjects barely cared about their co-player's payoff. Under the assumption that this also applies to the players in our group treatment, our findings indicate that group members lied together mainly for personal financial gain and not for helping their co-player. Only for religious players, we identify a strong willingness to engage in white lies. We explain this by the fact that players who felt more beneficent cared more about their co-players in the experiment and, therefore, were more likely to lie in order to raise their co-player's payoff (Erat & Gneezy, 2012; Gino et al., 2013).[31]

7.5. Conclusion

In general, people who cheat want to rationalize or downplay their behavior – and our paper investigates conditions that facilitate or impede this. For example, honesty oaths seem to make it harder to rationalize individual cheating behavior. In addition, our findings suggest that people rationalize their lying behavior by referring to the norm of helping others more often in two-player groups (in which both players benefit from telling a joint lie) rather than in a reciprocal relationship (in which liars do not benefit from lying themselves). By extension, this could imply that cheating behavior can be mitigated by making the rationalization of cheating more difficult or by providing people with less desired

[31] Ashforth and Anand (2003) correctly point out that corrupt systems and individuals are mutually reinforcing and that individuals joining a corrupt firm could be quickly indoctrinated by the "business as usual" mentality into the corrupt system. The firm dynamics aspects of institutionalization and socialization, discussed by Ashforth and Anand (2003), cannot be directly accounted for in any simple laboratory experiment.

rationalization opportunities. Our experimental evidence is not in line with widespread expectations.[32] While one would hesitate to recommend abandoning the four-eyes principle and thus returning to individual decision making, our results demonstrate how strongly honest decision making can be undermined by undesired collaboration. We argue that the negative effect of divided moral costs in groups persists regardless of the distribution of responsibilities between group members. Even if players are instructed and independently paid to monitor their co-player, they tend to cooperate in cheating. However, our findings also support the conjecture that players hardly cared about their co-player's payoff. On the contrary, we find that our participants rarely lied for each other, even if they were in a reciprocal relationship. We conclude that group members who agree to cheat together collaborate on lying mostly for personal gain. This may be the reason why measures that aim to raise the decision maker's moral awareness (e.g., honesty oaths) so often outperform group-based measures (e.g., the four-eyes principle or monitoring) with respect to their power to foster ethical behavior.

7.6. Implications

Our experimental results show that group decision making can have a negative effect on corporate responsibility, even when monitoring is implemented. Furthermore, our findings illustrate the risk that key decision makers behave opportunistically when given the chance to increase their payoff by deception. Although the moral costs of lying to help others appear to be comparatively low, our findings suggest that people do not typically care enough for others to engage in deceptive behavior for other persons. Finally, our results strongly encourage companies to establish a business culture that makes the moral costs of lying salient, for instance, by moral reminders (Mazar et al., 2008), such as honesty oaths.

In our experiment, signing the following statement was sufficient to yield more moral behavior among our subjects:

> "*Hereby I do affirm that all the data I am about to provide regarding my actions during this experiment will be the absolute truth. I also do swear that all my actions during this experiment will be due to the principle of honesty and that I will not lie in order to enrich myself.*"

De Bruin (2016), who describes public and performative ceremonies for oath-taking, suggests that most oaths could be improved by the inclusion of the Golden Rule, i.e., the law of reciprocity preached in

[32] To check whether we are the only ones to find our results partly surprising, we asked a different subject pool (100 students from the University of Kassel) in an online questionnaire to estimate our treatment results. We incentivized their guesses: For each treatment, the answer that was closest to our actual result in the respective treatment received €10. Subjects of our online survey systematically overrated the effect of monitoring: Whereas people believed that monitoring would be able to mitigate lying (they expected an average payoff of €3.21 in the monitoring treatments vs. €3.46 in Baseline, Wilcoxon signed-rank test: $p = 0.077$), we do not find this effect in our experimental data (on average, we paid €3.78 in the pooled monitoring treatments; two-sided Mann-Whitney U test to compare the mean rank of expected and actual payoffs: $p < 0.001$). Furthermore, people underestimated lying in the group treatment (they expected an average payoff of €3.32, actually it was €4.14; $p < 0.001$) and in the baseline scenario (expected €3.46 vs. actual €3.67; $p = 0.040$). However, our survey respondents slightly but not significantly overestimated lying in the reciprocity treatments (expected €3.13 vs. actual €2.89; $p = 0.881$).

most religions: “*Do unto others as you would have them do unto you*” or a variant thereof. Since our laboratory environment is not suitable for reconstructing an adequate oath-taking ceremony, we suggest future field experiments to verify the effectiveness of different kinds of oaths as forms of business ethics management (de Bruin, 2016). This would allow analyzing the impact that the practical framework of oaths has on their effectiveness, including the impact of the organization culture, the legal and regulatory environment, and the design of practices and institutions (Boatright, 2013). For instance, to test the effectiveness of an oath, the oath could be introduced at different locations of a company (compared to no oath for randomly selected employees). In addition, the company could randomly use different formulations (types) of oaths to analyze their comparative effectiveness. In order to evaluate the effectiveness of the oaths, the company could use surveys (half a year after swearing the oath) with a randomized response technique (Warner, 1965). With this technique, staffs can respond truthfully to moral questions without the company being able to conclude from single answers whether particular respondents engaged in immoral behavior. Yet, the company can learn whether there has been misconduct on an aggregate level. By this means, testing the effectiveness of honesty oaths comparable to ours in real companies could be a decisive step to verify the transferability of our findings to the business world.

CHAPTER 3

Size matters! Lying and Mistrust in the Continuous Deception Game

Tobias Beck♣
Institute of Economics, University of Kassel, Nora Platiel-Straße 4, 34127 Kassel, Germany

Abstract. I present a novel experimental design to measure lying and mistrust as continuous variables on an individual level. My experiment is a sender-receiver game framed as an investment game. It features two players: firstly, an advisor with complete information (i.e., the sender) who is incentivized to lie about the true value of an optimal investment and, secondly, an investor with incomplete information (i.e., the receiver) who is incentivized to invest optimally and therefore must rely on the alleged optimum reported by the advisor. Due to its continuous message space, this experiment allows observing more differentiated behavior and therefore enables testing of more sophisticated theoretical predictions. I find that the senders lie by overstating the true value of the optimum to an average extent of about 148%, while the receivers suspect them to do so by only 56%. Moreover, my results indicate that the senders make strategic considerations about their potential to manipulate others when deciding about the sizes of their lies. However, I find that the size of the lie and the size of mistrust do not only matter from a strategic perspective but also have an impact on how people perceive their own behavior. Consistent with previous studies, my findings support the conjecture that lying costs increase with the size of the lie. Beyond that, I provide evidence for some endogenous preference for trust. Both players' behaviors and beliefs are consistent over time. In addition, my classification of both players' strategies is consistent with their self-assessment of their behavior within the experiment.

Keywords: Size of the lie, Size of mistrust, Honesty, Deception Game, Investments, Asymmetric information, Experimental design.

JEL classification: C91, D01, D82, G41.

Availability of data: The datasets generated and analyzed in this paper are available from the author of this dissertation on reasonable request.

Informed consent: Informed consent was obtained from all individual participants included in the study.

♣ Email: tobi.beck@gmx.net; ORCID: 0000-0002-4162-3648

1. Introduction

The measurement of (dis)honesty and (mis)trust is a challenge for researchers in various fields of study. In particular, in experimental economics, there is a wide range of experimental designs for that purpose. However, the experimental literature on that subject often considers both of these factors separately and, in many cases, limits players' decision making to binary choices. In fact, to the best of my knowledge, there is no previous experimental design that allows measuring both the size of the lie *and* the size of mistrust on continuous scales at the same time. However, the ambiguity that is usually linked to telling a lie (or believing it) demands richer message spaces for the measurement of both of these concepts. Moreover, it should be considered that, especially in practice, lying and mistrusting behavior are closely linked to one another. In most business areas honesty and trust are of uttermost importance, for instance, in the consulting industry. This sector has been flourishing for years now, with the number of consultants rapidly increasing. It is not uncommon that private investors, managers, or even public officials involve consultants into their financial decisions. Typically, these advisors have (or claim to have) superior information or understanding of the consulting project than their clients, which makes the clients highly dependent on their advisors. If a conflict of interest between the advisors and the clients arises, this leaves room for opportunistic behavior by the advisors.

While some honest advisors certainly are inclined to give sound advice to their clients, others seek to maximize their own profit. In fact, the list of investment advisors who gave misleading or false advice to their clients is long (e.g., Dimmock & Gerken, 2012; InvestmentNews, 2019; Securities and Exchange Commission, 2008, 2019). For instance, in 2011 employees of one of HSBC's subsidiaries advised and sold savings products to over 2,400 elderly clients with investment periods longer than their life expectancy (Belfast Telegraph, 2011). As a result, a number of clients with shorter life expectancies than the recommended five-year investment timeframe had to make withdrawals from their investments sooner than recommended. Another infamous example is Yun Soo Oh Park's "pump and dump" scheme, which was discovered in 1999. In several instances, Park advised his clients to purchase shares of stocks in which he had already invested without his clients' knowledge, planning to sell his shares into the buying flurry and subsequent price rise that followed his recommendations (Securities and Exchange Commission, 2001). By this means, he made over $1.1 million in a one-year period (Fowler et al., 2001).

What all of these cases have in common is that dishonest advisors benefited financially from their clients' misinvestments. To achieve this, they lied about some private information, while their clients trusted their advice to be true. In a straightforward approach to model this situation, there are, on the one hand, the advisors who might have a preference for honest behavior that is in conflict with their desire for financial gain. On the other hand, there are the clients who face a situation of insecurity regarding some form of investment and, therefore, have to decide how much they can rely on their advisors. In the face of the above examples, the need to examine the core of such honesty-trust relationships is self-evident.

With that in mind, in this paper, I introduce a novel experimental design that allows for analysis of lying and mistrust in a two-player relationship with asymmetric information. This new experiment stands out from other experiments, as it permits the measurement of dishonesty, mistrust, and players' beliefs about one another on comparable *continuous* scales. On this basis, it allows for observation of more differentiated decision making and therefore makes it possible to test more sophisticated theoretical predictions. Another advantage of this experiment is that it can easily be conducted with pen and paper, takes only a short period of time to complete, and is easily extendable.

My *Continuous Deception Game* (CDG) is inspired by Gneezy (2005) as well as Erat and Gneezy (2012). It is a sender-receiver game framed as an investment game. Thus, it features two players: in the first place, an advisor with complete information (i.e., the sender) who is incentivized to lie about the true value of an optimal investment and, in the second place, an investor with incomplete information (i.e., the receiver) who is incentivized to invest optimally but has no other information than the alleged optimum reported by the advisor. In order to minimize context effects, the game adds no specific context to the type of investment. However, practical cases of investments with a similar structure can easily be found: for instance, the decision on the optimal death benefit and the associated insurance premium one pays for their life insurance, or an optimal target contract sum of a construction loan.[33]

I contextualize my experimental design by providing an overview of related work in the literature review in section 2. Section 3 then describes the payoff structure, incentives, and game process of the CDG. This allows me to define several key variables to measure lying, mistrust, and players' expectations about each other in the game. After that, I categorize all feasible strategies in the game and show which of them are rational from a game theoretical perspective. On this basis, I formulate my hypotheses in section 4. I proceed by explaining the precise implementation of my experiment by discussing the design in section 5 and the experimental procedures in section 6. I then report the results in section 7. This section is divided into four subsections: (*7.1.*) the main results on both players' behavior and first-order beliefs, (*7.2.*) an analysis of both players' strategies based on the relationship between their behavior and first-order beliefs, (*7.3.*) additional results on players' strategic considerations about the potential to manipulate others in the game, and lastly (*7.4.*) a short summary of key results. In section 8, I then discuss my results in the light of the existing literature. Finally, section 9 concludes my most important findings.

[33] Other examples include a company's decision about the optimal size of a new industrial facility, as well as informative lobbying where the government relies on lobbyists who may have superior information that could help to make better-informed policy choices.

2. Literature review

2.1. Honesty

There is broad evidence that people have some form of preference for honest behavior (e.g., Charness & Dufwenberg, 2006; Erat & Gneezy, 2012; Fischbacher & Föllmi-Heusi, 2013; Gneezy, 2005; Gneezy et al., 2018; Hurkens & Kartik, 2009; López-Pérez & Spiegelman, 2013; Lundquist et al., 2009; Mazar et al., 2008; Vanberg, 2008). In general, it is argued that this preference is intrinsic rather than extrinsic. In support of this, Pruckner and Sausgruber (2013) show that appealing to honesty can mitigate dishonest behavior more effectively than a reminder of legal norms. Often this is at least partially explained by the concept of lie aversion (e.g., Fischbacher & Föllmi-Heusi, 2013; Gneezy, 2005; Gneezy et al., 2018; Hurkens & Kartik, 2009; López-Pérez & Spiegelman, 2013; Lundquist et al., 2009). Vanberg (2008) even provides experimental evidence that people dislike the act of lying per se. Another approach comes from Charness and Dufwenberg (2006) who show that people's preference for honest behavior in their experiment can be explained by guilt aversion. On that subject, Battigalli et al. (2013) argue that, in some situations, guilt can provide a psycho-foundation for honesty. Moreover, Gneezy et al. (2018) find that the social identity of a person can influence this person's honesty. In fact, honesty seems to concern one's very identity. For instance, Blok (2013) argues that taking an oath changes the oath-taker's identity by manifesting their intention "not only to *do* something, but also to *be* the one who is committed to some future course of action" (Blok, 2013, p. 193, italics in original). In line with this, Mazar et al. (2008) find that directing one's attention to moral standards can lower one's tolerance for dishonest behavior. Here, Mazar et al. (2008) suggest that people who face a trade-off between some financial benefit from cheating and maintaining a positive self-concept try to find a balance between these two motivational forces. This indicates that individual preferences are a combination of selfishness and morality, which implies a *homo moralis*-like conception of man (Alger & Weibull, 2013).

In many studies, these internal motivators for preferences for honesty are modeled by the concept of costs of lying (e.g., Beck et al., 2020; Gneezy et al., 2018; Lundquist et al., 2009). Therefore, it is assumed that people assign a negative value to dishonest behavior for one or several of the above reasons. On that matter, Lundquist et al. (2009) find that the aversion to lying increases with the *size of the lie*. Following this idea, Beck et al. (2020) introduce a straightforward model of the utility of lying that includes lying costs that depend on the size of the lie. This model is able to predict honesty in several variations of Fischbacher and Föllmi-Heusi's (2013) dice experiment. In addition, Gneezy et al. (2018) present intrinsic costs of lying as a concept that is connected to different dimensions of the size of the lie. They distinguish between an outcome dimension (i.e., the difference between a reported value and the truth), a payoff dimension (i.e., the monetary gains from lying), and a likelihood dimension that reflects concerns about one's social identity (i.e., one's concerns about how one is perceived by others).

The range of motivators for honest behavior demonstrates that there are various reasons why one could choose *not* to lie. This intrinsic preference for honesty can be modeled by costs of lying, which seem to be systematically connected to the extent of the lie.

2.2. Trust

Honesty is closely related to trust. Trust, in turn, is important for various reasons. One reason is that social trust can raise economic growth rates (Beugelsdijk et al., 2004; Bjørnskov, 2012; Knack & Keefer, 1997). Trust is important for the banking sector in particular (Boatright, 2013). Also, with respect to the individual, Barefoot et al. (1998) show that high levels of trust on Rotter's (1967) interpersonal trust scale are associated with better self-rated health and more life satisfaction. In addition, Kuroki (2011) finds, by analyzing survey data, that interpersonal trust has positive and significant effects on individual happiness. In line with this, Gurtman (1992) provides evidence that extreme distrust in interpersonal relationships is related to distress. This provides reason to assume that people might have an intrinsic preference for trust. However, since other people are not necessarily trustworthy, this preference might conflict with a preference for risk aversion. This is supported by Sapienza et al. (2013) who find that the quantity sent in the Trust Game depends not only on the trustor's belief in the other player's trustworthiness but also on the trustor's risk aversion.

In general, having doubts about another person's trustworthiness is closely related to the beliefs one has towards this person. According to McKnight and Chervany (2001, p. 36), "trusting beliefs are cognitive perceptions about the attributes or characteristics of the trustee". If an individual believes someone to be trustworthy, they can build the intention to trust that person and eventually treat that person with trusting behavior. This distinction is in line with McKnight and Chervany's (2001) constructs of trusting beliefs, intention, and behavior. Beyond that, trust can be directed towards different traits of another person, such as a person's honesty or degree of social cooperation. In this paper, I am interested in the relationship between honesty and trust. Therefore, "trust" here refers to *honesty-related trusting behavior*, i.e., one's reliance on another person's honesty, whereas "trusting beliefs" represent expectations about the honesty of another person.

Which type of trust is observed in experimental studies depends on the experimental design. One of the most famous games that aim to model trust is the aforementioned Trust Game. In fact, many experimental studies that examine means to enhance trust (e.g., promises, oaths or gifts) are based on variations of the Trust Game (e.g., Charness & Dufwenberg, 2006, 2010; Ismayilov & Potters, 2016; Servátka et al., 2011). In the original version of this game, the trustee's choice on how much to send back to their trustors depends primarily on their preferences for social cooperation and not for honesty, since not sending back anything might be unsocial but not dishonest. As a consequence, "trust" in the original form of the Trust Game refers to the act of relying on another person's social cooperation but not on another person's honesty. This example shows that in order to observe honesty-related trust it is important to make sure that the trustor depends on the honesty of the trustee. For this reason, the trustor needs to be given a task for the fulfillment of which he or she has to decide whether to trust or mistrust the trustee. This in turn requires that the trustee has some information advantage over the trustor. A game that meets these requirements is the sender-receiver game, which I will focus on in the next subsection.

It can be concluded that people seem to have a general preference for trust. Their trust, however, is in conflict with their individual risk aversion and concerns about the other person's trustworthiness. Moreover, modeling honesty-related mistrust requires some degree of asymmetric information between the trustor and the trustee.

2.3. Modeling honesty and trust in games with incomplete information

Information asymmetries are of the uttermost importance for strategic decision making. There are many examples of people suffering from incomplete information in the business world. For instance, in the finance sector, managers normally have better information about their firms than their shareholders (Boatright, 2013; Sobel, 2009). The same applies to advisors who have more information than the investors who consult them. In general, insiders have superior information to investors (Leland & Pyle, 1977). Hence, insiders might use their informational power to manipulate investors to invest in their firm (Sobel, 2009). In all of these situations, one party with more information could use their information advantage to exploit another party with less information.

According to Sobel (2009), these problematic situations can be adequately modeled by designing appropriate *sender-receiver games*.[34] He narrowly defines such games as a class of two-player games of incomplete information. What makes this type of experiment so suitable for examining honesty and trust at the same time is that it defines clear strategy sets for the informed and the uninformed player (Sobel, 2009), which determine honest behavior, on the one hand, and trusting behavior, on the other hand. Here, the informed player is typically referred to as "sender" and the uninformed player as "receiver".

Experiments based on this type of game are widely used in order to analyze (dis)honest and (mis)trusting behavior. For instance, Gneezy (2005), Sánchez-Pagés and Vorsatz (2007), Dreber and Johannesson (2008), Sutter (2009), Erat and Gneezy (2012), López-Pérez and Spiegelman (2013), Peeters et al. (2013), Peeters et al. (2015), Jacquemet et al. (2018), Jacquemet, Luchini, Rosaz, et al. (2019), Vranceanu and Dubart (2019), and Gneezy et al. (2020) implement versions of sender-receiver games in their studies – to name only a few.

As this paper is largely inspired by Erat and Gneezy's (2012) *Deception Game* (which is a sender-receiver game that originated from Gneezy, 2005), I will now discuss their experimental design in more

[34] For a more detailed discussion of the question of whether or not dishonest behavior such as corruption can be studied in the laboratory, see Armantier and Boly (2008).

detail. Their game begins with the sender being informed about the outcome of a roll of a six-sided die. Then, he or she is asked to communicate the outcome of the die roll to the receiver by choosing from a pool of six possible messages. There is one message for each possible outcome of the die roll, each stating that "the outcome from the roll of the 6-sided die is…" (Erat & Gneezy, 2012, p. 731) the corresponding number between 1 and 6. After receiving this message, the receiver has to choose a number between 1 and 6. This choice determines which of two payoff options, A or B, is implemented. Here, it is known to both players that if the receiver chooses the actual outcome of the die roll, option A is implemented, and if he or she chooses any other number, option B is implemented. However, only the sender is informed about the payoffs associated with both payoff options. This gives the sender the opportunity to lie in order to manipulate the receiver into picking a number associated with payoff option B. Erat and Gneezy (2012) use this mechanism by manipulating the change in payoffs between both payoff options in order to implement treatments with different types of lies. On this basis, they distinguish between altruistic white lies (i.e., lies that are expected to reduce the sender's payoff while increasing that of the receiver), Pareto white lies (i.e., lies that are expected to increase both players' payoffs), spiteful black lies (i.e., lies that are expected to reduce both players' payoffs), and selfish black lies (i.e., lies that are expected to increase the sender's payoff while reducing that of the receiver).[35]

In Erat and Gneezy's (2012) experiment – as well as in all other above mentioned versions of sender-receiver games – both players are confronted with discrete (or in most cases even binary) choices. In the case of Erat and Gneezy's (2012) experiment, both players' decisions can be considered as *binary-like* choices.[36] Thus, honesty and trust are measured as dichotomous variables. With regard to the sender, honest behavior can solely be distinguished from one single predefined type of lying. This is because the sender cannot choose in which of the different types of lies he or she wants to engage, as the only change in payoffs that he or she can achieve by deceiving the receiver is predefined by the experimenters.[37] As for the receiver, the Deception Game permits distinguishing solely trusting behavior from mistrusting behavior. While these simplifications serve the purpose of Erat and Gneezy's (2012) paper in a good way, in reality the ranges of dishonesty and mistrust are more continuous.

Lundquist et al. (2009) deal with this matter by implementing a sender-receiver game that allows for different sizes of lies but again features a binary payoff structure for the senders. Their game features a seller who can lie about their talent and a buyer who can send the seller a fixed-payment contract. If the

[35] A similar distinction is also used by Gneezy (2005).
[36] Note that both players can choose between six different options (related to numbers between 1 and 6). It is reasonable to assume, however, that players have no preferences for specific numbers beyond preferences that could arise due to the rules of the game. Thus, the receiver's trust should be independent of the number that the sender communicates to them. For this reason, the receiver should be indifferent between the five numbers that were not communicated to them by the sender. If the sender anticipates this, he or she should also be indifferent between the five untruthful messages that he or she can send.
[37] Translated to Gneezy et al.'s (2018) model of intrinsic costs of lying, Erat and Gneezy's (2012) taxonomy of lies is based on the payoff dimension of the size of the lie. Here, the payoff dimension can still be interpreted on a continuous scale. However, since the sender can only choose between honest behavior and one single predefined type of lying, a single decision of one sender captures this dimension as a dichotomous variable.

contract is signed, the buyer makes a loss if the seller's talent is below a certain threshold and a profit otherwise. Here, the seller's talent is defined by a given number between 1 and 100. Moreover, the payment of the seller is higher if the contract is signed. Therefore, a seller with a talent score below the given threshold has an incentive to lie about their talent to ensure a contract. In this experiment, the extent to which the seller needs to lie in order to achieve a contract depends on the difference between the seller's true talent and the given threshold, which are both predefined by the experimenters.[38] The resulting binary payoff structure of the sender greatly limits the possibilities of comparing two lies with different sizes to one another if both lies are expected to result in the same payoff. Hence, even though the sender's choice is continuous, their behavior can barely be interpreted on a continuous scale.[39]

It can be concluded that any experiment that measures honesty and trust as dichotomous variables falls short of observing which type of lying or mistrusting behavior the players actually *do* prefer. For instance, such experiments cannot reveal whether or not a player who told a Pareto white lie would rather have preferred to tell a selfish black lie (or any other type of lie) if given the choice. Since binary choice-based experiments cannot address this matter on an individual level, results of sender-receiver games usually report the relative frequencies of observed lies and mistrust on a group level. These frequencies do not represent feasible strategies in the game but proportions of senders or receivers who lied or, respectively, mistrusted their co-players to a predefined extent. Therefore, such frequencies do not provide any information on the extents of lying or mistrust and are unsuitable for capturing how dishonest or mistrusting single players behaved.[40] As a consequence, binary (or even discrete) choices do not allow measuring the real extent to which the sender (or the receiver) would prefer to lie (or, respectively, to mistrust) when given the chance.

To the best of my knowledge, no previous experiment allows examining honesty *and* trust as continuous variables on comparable scales. This is where this paper aims to contribute.

3. The Continuous Deception Game

In order to analyze the relationship between players' dishonesty, mistrust, and their expectations of each other's behavior, I designed a novel experiment. It is a complex version of a sender-receiver game

[38] Lundquist et al. (2009) determined the talent score of the sellers based on a test that took place before the actual experiment started.

[39] With respect to Gneezy et al.'s (2018) model of intrinsic costs of lying, Lundquist et al.'s (2009) experimental design allows measuring only the outcome dimension of the size of the lie on a continuous scale. The payoff dimension, however, is reduced to a dichotomous variable, which makes it difficult to interpret the size of the lie as continuous.

[40] Note that there are other experimental designs that allow the players to choose between different types or degrees of lying. One famous example is the dice experiment of Fischbacher and Föllmi-Heusi (2013), which allows distinguishing partial from payoff-maximizing lies. Another design is the experimental design of Gneezy et al. (2018) who extend this idea by introducing an *n*-sided die to this form of cheating game. The die in their experiment is implemented via computerization as well as by using an envelope with *n* folded pieces of paper that have numbers from one to *n* written on them. However, to the best of my knowledge, not a single experimental design that aims to measure the size of lying also allows observing trust on a comparable scale.

(as defined by Sobel, 2009) that is framed as an investment game.[41] Therefore, it features an advisor with complete and an investor with incomplete information. This novel game is largely inspired by the experimental designs of Gneezy (2005) and Erat and Gneezy (2012). It expands Erat and Gneezy's (2012) Deception Game by introducing continuous strategy sets for dishonesty and mistrust. This allows measuring these two variables on easily comparable continuous scales and, thus, observing more differentiated decision making. For this reason, I name my experiment the *Continuous Deception Game* (CDG).

3.1. *Payoff structure and incentives*

Towards the end of the CDG, the investor has to make an investment by choosing any number i between 0 and a predefined maximum $i_{max} > 0$. The investment i then determines the individual payoff of both players. There is one unique optimal investment $i^* \in [0, i_{max}]$ that maximizes the investor's payoff and is randomly determined by nature before the game starts.[42] Both players have different payoff functions:

In the first place, by design the *advisor's payoff* increases with the investment i. For simplicity, the advisor's payoff $\pi_A(i)$ is defined as a linearly increasing function of the investment i:

$$\pi_A(i) = m_A * i \tag{6}$$

with $m_A > 0$.[43]

Note that the advisor's payoff is fully dependent on the investor's behavior. In particular, he or she receives nothing if the investor chooses not to invest ($i = 0$), whereas he or she maximizes their payoff if the investor chooses to make the maximal investment ($i = i_{max}$). Thus, if the optimal investment i^* is not equal to the investment's maximum i_{max}, the advisor is monetarily incentivized to get their investor to make an overinvestment (for $i^* \neq i_{max}$: $i > i^*$).

In the second place, the *investor's payoff* is designed to decrease by any downward or upward deviation of the investment i from its optimum i^*. In order to be able to make any investment i the investor starts with an initial amount, which is equal to the maximal investment i_{max}. If the investor decides not to make an investment ($i = 0$), he or she keeps their initial amount, resulting in a payoff $\pi_I(i)$ equal to the maximal investment i_{max} (if $i = 0$: $\pi_I = i_{max}$). To meet these conditions, the investor's payoff $\pi_I(i)$

[41] The idea of framing my experiment as an investment game is inspired by Berg et al. (1995) who designed an investment game in order to introduce continuous variables to the Trust Game.

[42] The optimal investment i^* is randomly determined by a uniform distribution: $P([0, i^*]) = \frac{i^*}{i_{max}}$ with $i^* \in [0, i_{max}]$.

[43] The factor m_A is the advisor's payoff factor. This factor determines to which extent he or she profits from the investment i.

is defined by the following split function of the investment i, which decreases linearly by deviating downward or upward from the optimal investment i^*:

$$\pi_I(i) = \begin{cases} i \leq i^*: & i_{max} + m_I * i \\ i > i^*: & i_{max} + m_I * (2 * i^* - i) \end{cases} \tag{7}$$

with $m_I > 0$.[44]

It should be borne in mind that the investor is monetarily incentivized to try and make an optimal investment ($i = i^*$), since this maximizes their payoff. Moreover, the higher the investment i, the more the investor could lose from (or win on top of) their initial amount i_{max}. Thus, the lower the investment i, the lower is the financial risk for the investor.

Figure 3-1 shows an example of both players' payoff functions $\pi_A(i)$ and $\pi_I(i)$. Here, the maximal investment i_{max} is assumed to be equal to 100, while the optimal investment i^* is 50. In addition, the payoff factors m_A and m_I are assumed to be equal to 0.5.

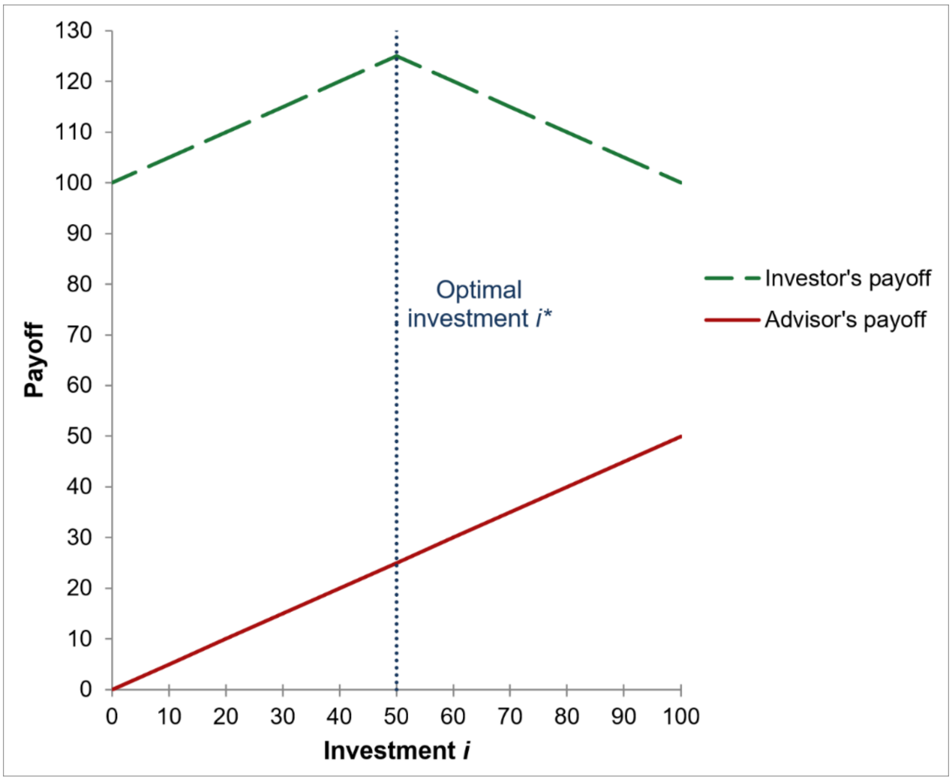


Figure 3-1. Payoff structure of the CDG

Obviously, the advisor and the investor have conflicting monetary incentives. From a financial perspective, the investor wants to try to make an optimal investment ($i = i^*$), whereas the advisor prefers the investment to be as high as possible ($i = i_{max}$). If the optimum i^* is not equal to the maximal

[44] The factor m_I is the investor's payoff factor. This factor determines to which extent he or she can profit or lose from their investment i. Note that whether the investor profits or loses from making an investment i also depends on the value of the optimal investment i^*.

investment i_{max}, this would mean that the advisor would want the investor to overinvest (for $i^* \neq i_{max}$: $i > i^*$). These conditions are common knowledge among the players.

3.2. *The game process*

Figure 3-2 summarizes the process of the CDG. The game itself consists of two main stages. As mentioned before, the optimal investment i^* is determined randomly and in secret by nature before the game starts. Then, in the first stage of the game, only the advisor is informed about the value of the optimal investment i^*. Afterwards, he or she is instructed to report this optimum to the investor. To this end, the advisor chooses an advice number $a \in [0, i_{max}]$ that will be sent to the investor later as advice on the value of the optimal investment i^*. Thus, the advice a can be considered as completely truthful if it is equal to the optimal investment i^* ($a = i^*$). In addition, the advisor is asked to make a guess $i_{guess} \in [0, i_{max}]$ which investment i the investor will make later based on their advice a.[45] This guess i_{guess} reflects the advisor's first-order beliefs about the investor. Note that if this guess i_{guess} is equal to the given advice a ($i_{guess} = a$), the advisor thereby states that he or she expects their investor to exactly follow their advice a. This would mean that the advisor believes the investor will behave with complete trust.

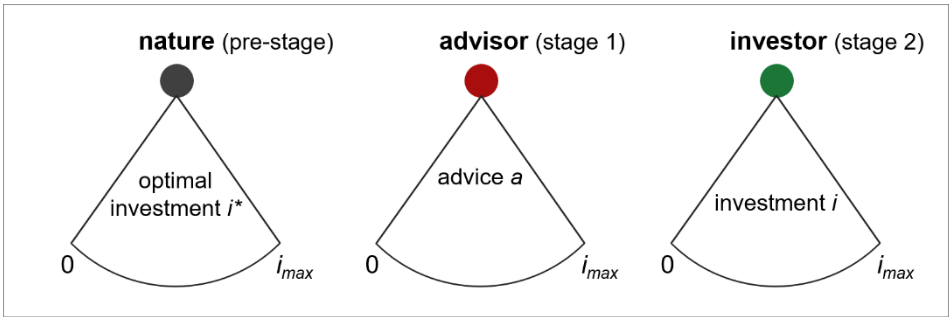


Figure 3-2. Game tree of the CDG

In the second stage of the game, the investor is informed about the advice a. Then he or she is instructed to make their investment $i \in [0, i_{max}]$. Hence, the behavior of the investor can be considered as completely trusting if he or she makes an investment i equal to the received advice a ($i = a$). Apart from that, the investor is asked to make a guess $i^*_{guess} \in [0, i_{max}]$ which might be the true optimal investment i^*.[46] This guess i^*_{guess} reflects the investor's first-order beliefs about the advisor. Note that

[45] This is inspired by Sutter (2009) who also asked the senders in his sender-receiver game which response they would expect from the receivers in order to observe the intention behind their behavior. Similar to Sutter (2009), I refrain from monetarily incentivizing this guess because, according to Sutter (2009) who refers to Camerer and Hogarth (1999), "there is evidence that eliciting expectations with or without monetary rewards for accuracy does not yield significantly different results" (Sutter, 2009, p. 50).

[46] This guess allows analyzing whether the investor tries to maximize their payoff (which would imply: $i = i^*_{guess}$). As for the advisor, the investor's guess is not monetarily incentivized, since this is not expected to significantly change the quality of the guess (Camerer & Hogarth, 1999; Sutter, 2009).

if this guess i^*_{guess} is equal to the received advice a ($i^*_{guess} = a$), the investor expects the advice a to be truthful. This would mean that the investor does not suspect their advisor to have lied.

Subsequently, the game ends and the payoffs of both players are determined by the investment i according to their payoff functions $\pi_A(i)$ and $\pi_I(i)$ as defined above. Due to the continuous message space of the CDG (and in contrast to binary choice-based sender-receiver games, such as those of Gneezy, 2005; López-Pérez & Spiegelman, 2013; Peeters et al., 2015; Sánchez-Pagés & Vorsatz, 2007), this design can address the issue of sophisticated deception through truth-telling (Sutter, 2009).[47]

3.3. Key variables

The CDG allows measuring seven key variables, which I will define in the following subsections.

3.3.1. (Suspected) lying

Definition 1. The percentage *extent of lying* L of the advisor is defined as:

$$L = \frac{a - i^*}{i^*} \tag{8}$$

with $i^* > 0$.[48]

Note. A piece of advice a can be considered as *truthful* ($L = 0$) if it is equal to the true optimal investment i^* (if $a = i^*$: $L = 0$). All other pieces of advice can be considered as *lies* (if $a \neq i^*$: $L \neq 0$). In particular, lying by giving advice a below the true optimum i^* is defined as *lying by understating* (if $a < i^*$: $L < 0$), whereas giving advice a above the optimal investment i^* is considered as *lying by overstating* (if $a > i^*$: $L > 0$).

[47] Sophisticated deception through truth-telling refers to cases in which the senders expect their receivers to mistrust them (by not following their message) and send the true message for precisely this reason (Sutter, 2009).

[48] The extent of lying is expressed as a percentage here because, without any further reference point, the absolute deviation of the advice a from the optimal investment i^* (e.g., $a - i^* = 1$) would not properly reflect the gravity of the lie. For instance, lying by overstating the optimal investment by 1 can be considered as more dishonest if $i^* = 1$ and $a = 2$ rather than $i^* = 41$ and $a = 42$. To deal with this issue, I defined the extent of lying in proportion to the optimal investment i^*, which constitutes the reference point for "truth-telling" in the CDG.
Note that another intuitive way to define the percentage extent of lying would be to refer to the effect the lie has on the investor's payoff $\pi_I(i)$ if he or she follows the received advice a. Following this idea, the extent of lying could also be defined as the percentage extent to which the investor's payoff $\pi_I(i)$ would be reduced by following the advice a when compared to making an optimal investment i^*, which would imply: $\frac{\pi_I(i^*) - \pi_I(a)}{\pi_I(i^*)}$. However, using this alternative definition for the extent of lying does not change any of the general results presented in this paper. Therefore, and since this definition would be incoherent with the definitions of the other key variables, I decided in favor of the initially presented definition.

Definition 2. The percentage *extent of suspected lying* L_{guess} to which the investor suspects their advisor to lie is defined as:

$$L_{guess} = \frac{a - i^*_{guess}}{i^*_{guess}} \tag{9}$$

with $i^*_{guess} > 0$.

Note. This extent reflects the investor's trusting beliefs, which correspond to their expectations about the advisor's honesty. If the investor's guess about the optimal investment i^*_{guess} is equal to their received advice a, he or she thereby considers the advice a to be *truthful* (if $a = i^*_{guess}$: $L_{guess} = 0$). Thus, in all other cases the investor suspects their advisor to have *lied* (if $a \neq i^*_{guess}$: $L_{guess} \neq 0$). More precisely, guessing that the optimal investment i^*_{guess} is above the advice a shall be defined as *suspecting an understating lie* (if $a < i^*_{guess}$: $L_{guess} < 0$), whereas guessing that the optimal investment i^*_{guess} is below the advice a can be considered as *suspecting an overstating lie* (if $a > i^*_{guess}$: $L_{guess} > 0$).

3.3.2. (Expected) mistrust

Definition 3. The percentage *extent of mistrust* $\bar{T}$ to which the investment i deviates from the received advice a is defined as:

$$\bar{T} = \frac{i - a}{a} \tag{10}$$

with $a > 0$.

Note. In the CDG, trust refers to the investor's trusting behavior, i.e., the investor's reliance on their received advice number. This definition of trust is in line with previous sender-receiver games (e.g., Peeters et al., 2013; Sutter, 2009), in which trust is defined as the act of following the message from the sender. With that in mind, the behavior of the investor can be considered as completely *trusting* ($\bar{T} = 0$) if their investment i is equal to their received advice a (if $i = a$: $\bar{T} = 0$). As a result, all other behaviors shall be defined as *mistrusting* (if $i \neq a$: $\bar{T} \neq 0$). In particular, making an investment i below the received advice a is considered as *risk-reducing mistrust* (if $i < a$: $\bar{T} < 0$), while making an investment i above the advice a is defined as *risk-seeking mistrust* (if $i > a$: $\bar{T} > 0$), since in this game higher investments i are associated with a higher risk.[49]

[49] The distinction between risk-reducing and risk-seeking mistrust is in line with Sapienza et al. (2013) who find that mistrust is associated with a preference for risk aversion in the Trust Game. Another reason for using this terminology is that, in the CDG, the possible variance of the investor's payoff increases with the size of the investment i. This is because the higher the investment i, the more the investor can win or lose from their investment by design.

Definition 4. The percentage *extent of expected mistrust* $\bar{T}_{guess}$ to which the advisor expects the investment i to deviate from their given advice a is defined as:

$$\bar{T}_{guess} = \frac{i_{guess} - a}{a} \tag{11}$$

with $a > 0$.

Note. This extent reflects the advisor's beliefs about the investor's trusting behavior. If the advisor's guess about the investment i_{guess} equals their given advice a, he or she thereby considers the investor to behave completely *trusting* (if $i_{guess} = a$: $\bar{T}_{guess} = 0$). In all other cases the advisor states that he or she expects their investor to behave *mistrusting* (if $i_{guess} \neq a$: $\bar{T}_{guess} \neq 0$). To specify, guessing that the investment i_{guess} is below the given advice a shall be defined as *expecting risk-reducing mistrust* (if $i_{guess} < a$: $\bar{T}_{guess} < 0$), whereas guessing that the investment i_{guess} is above the advice a shall be considered as *expecting risk-seeking mistrust* (if $i_{guess} > a$: $\bar{T}_{guess} > 0$). Again, this terminology is based on the fact that in this game higher investments i are associated with a higher risk by design.

3.3.3. (Expected) misinvestment

Definition 5. The *advisor's* percentage *extent of expected misinvestment* $F_{A;guess}$ to which he or she expects the investment i_{guess} to deviate from its optimum i^* is defined as:

$$F_{A;guess} = \frac{i_{guess} - i^*}{i^*} \tag{12}$$

with $i^* > 0$.

Note. If the advisor's guess about the investment i_{guess} is equal to the optimal investment i^*, the advisor thereby states that he or she expects the investment to be *optimal* (if $i_{guess} = i^*$: $F_{A;guess} = 0$). In all other cases the advisor expects a *misinvestment* to some extent (if $i_{guess} \neq i^*$: $F_{A;guess} \neq 0$). More precisely, guessing that the investment i_{guess} is below the true optimum i^* can be defined as *expecting an underinvestment* (if $i_{guess} < i^*$: $F_{A;guess} < 0$), while guessing that the investment i_{guess} is above the optimal investment i^* shall be considered as *expecting an overinvestment* (if $i_{guess} > i^*$: $F_{A;guess} > 0$).

Definition 6. The *investor's* percentage *extent of expected misinvestment* $F_{I;guess}$ to which he or she expects their investment i to deviate from their guessed optimal investment i^*_{guess} is defined as:

$$F_{I;guess} = \frac{i - i^*_{guess}}{i^*_{guess}} \tag{13}$$

with $i^*_{guess} > 0$.

Note. If the investment i is equal to the investor's guess about the optimal investment i^*_{guess}, he or she thereby considers the investment to be *optimal* (if $i = i^*_{guess}$: $F_{I;guess} = 0$). In all other cases the investor expects making a *misinvestment* to some extent (if $i \neq i^*_{guess}$: $F_{I;guess} \neq 0$). In particular, guessing that the investment i is below the guessed optimal investment i^*_{guess} shall be considered as *expecting an underinvestment* (if $i < i^*_{guess}$: $F_{I;guess} < 0$), whereas guessing that the investment i is above the estimated optimum i^*_{guess} is defined as *expecting an overinvestment* (if $i > i^*_{guess}$: $F_{I;guess} > 0$).

Definition 7. The actual percentage *extent of misinvestment* F to which the investment i deviates from its optimum i^* is defined as:

$$F = \frac{i - i^*}{i^*} \tag{14}$$

with $i^* > 0$.

Note. If the investment i is equal to the optimal investment i^*, it is considered as *optimal* by design (if $i = i^*$: $F = 0$). In all other cases it shall be defined as a *misinvestment* to some extent (if $i \neq i^*$: $F \neq 0$). To specify, if the investment i is below the optimal investment i^*, it is an *underinvestment* (if $i < i^*$: $F < 0$), whereas if the investment i is above its optimum i^*, it is an *overinvestment* (if $i > i^*$: $F > 0$).

This experiment allows measuring both the extent of lying by the sender, i.e., the advisor, and the extent of mistrust by the receiver, i.e., the investor (*see Definitions 1 and 3*). At the same time, it permits the measurement of both players' first-order beliefs, i.e., their expectations of their co-player's behavior (*see Definitions 2 and 4*). In addition, it allows measuring the quality of the outcome of a task with contradicting incentives, i.e., the investment, (*see Definition 7*) as well as both players' expectations towards it (*see Definitions 5 and 6*).

3.4. Taxonomy of feasible strategies

In the CDG, there are a great variety of strategies that can be pursued by both players. For that reason, it makes sense to define classes of feasible strategies for the advisor and the investor. To distinguish different types of strategies, I use the previously defined key variables, which describe both players' behavior and expectations. I will begin with the advisor (*3.4.1.*) and then turn to the investor (*3.4.2.*).

3.4.1. Taxonomy of lies and truth-telling

Figure 3-3 gives an overview of my taxonomy of lies and truth-telling for the advisor based on their percentage extent of lying (L on the ordinate) and their percentage extent of expected mistrust ($\bar{\bar{T}}_{guess}$ on the abscissa). In addition, the figure illustrates which financial outcome the advisor expects from different combinations of lying behavior and expected mistrust. This is done by taking the advisor's expectation about the extent of misinvestment ($F_{A;guess}$) into account. For this purpose, the dotted line in the figure represents strategies in which the advisor expects the investment to be optimal

($F_{A;guess} = 0$). Thus, an advisor with a strategy below this line expects the investor to make an underinvestment ($F_{A;guess} < 0$), whereas an advisor with a strategy above it expects an overinvestment ($F_{A;guess} > 0$). These expectations can be used to determine the change in both players' payoffs that the advisor anticipates due to their pursued strategy. Inspired by Erat and Gneezy (2012) who use the expected change in payoffs in order to distinguish different types of lies,[50] I define classes of feasible strategies for the advisor based on the combination of their lying behavior (L) and their expectations towards both players' payoffs (which are reflected in their expectations towards the outcome of the investment $F_{A;guess}$).

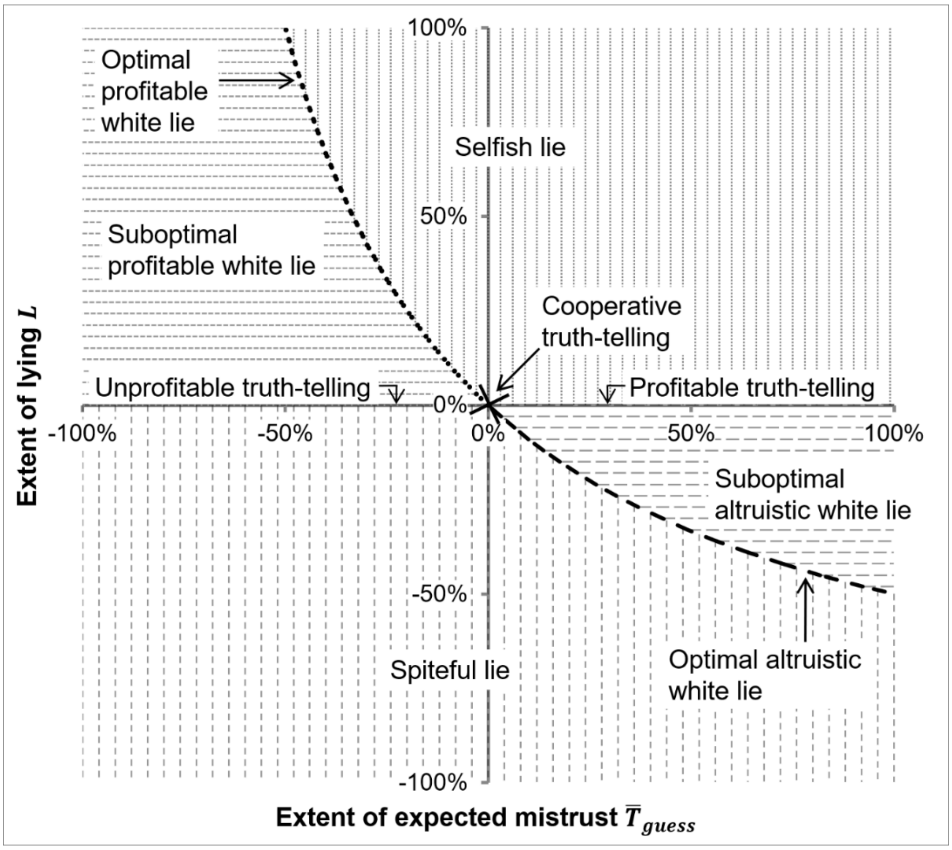


Figure 3-3. Taxonomy of lies and truth-telling for the advisor

As summarized in Table 3-1, there are nine classes of feasible strategies for the advisor:

Definition 8. *Spiteful lie*: The advisor lies by understating ($L < 0$) and therefore expects an underinvestment ($F_{A;guess} < 0$). Compared to more honest behavior, the advisor would expect this to reduce both players' payoffs.

Definition 9. *Optimal altruistic white lie*: The advisor lies by understating ($L < 0$) while expecting an equal extent of risk-seeking mistrust from the investor. As a result, the advisor expects the investment to be optimal ($F_{A;guess} = 0$).

[50] The four types of lies defined by Erat and Gneezy (2012) are described in section 2.3. above.

Definition 10. *Suboptimal altruistic white lie*: The advisor lies by understating ($L < 0$) while expecting an even stronger extent of risk-seeking mistrust from the investor. Hence, the advisor expects an overinvestment ($F_{A;guess} > 0$).

Definition 11. *Unprofitable truth-telling*: The advisor gives truthful advice ($L = 0$) but expects risk-reducing mistrust from the investor. Thus, the advisor expects an underinvestment ($F_{A;guess} < 0$).

Definition 12. *Cooperative truth-telling*: The advisor gives truthful advice ($L = 0$) believing that the investor will trust them. As a consequence, the advisor expects the investment to be optimal ($F_{A;guess} = 0$).

Definition 13. *Profitable truth-telling*: The advisor gives truthful advice ($L = 0$) but expects risk-seeking mistrust from the investor. Thus, the advisor expects an overinvestment ($F_{A;guess} > 0$).

Definition 14. *Suboptimal profitable white lie*: The advisor lies by overstating ($L > 0$) while expecting an even stronger extent of risk-reducing mistrust from the investor. As a result, the advisor expects an underinvestment ($F_{A;guess} < 0$).

Definition 15. *Optimal profitable white lie*: The advisor lies by overstating ($L > 0$) while expecting an equal extent of risk-reducing mistrust from the investor. Hence, the advisor expects the investment to be optimal ($F_{A;guess} = 0$).

Definition 16. *Selfish lie*: The advisor lies by overstating ($L > 0$) and therefore expects an overinvestment ($F_{A;guess} > 0$). Compared to more honest behavior, the advisor would expect this to increase their payoff while reducing that of the investor.

		Lying behavior		
		Understating lie ($L < 0$)	Truth-telling ($L = 0$)	Overstating lie ($L > 0$)
Expected investment	Underinvestment ($F_{A;guess} < 0$)	*Spiteful lie*	*Unprofitable truth-telling*	*Suboptimal profitable white lie*
	Optimal investment ($F_{A;guess} = 0$)	*Optimal altruistic white lie*	*Cooperative truth-telling*	*Optimal profitable white lie*
	Overinvestment ($F_{A;guess} > 0$)	*Suboptimal altruistic white lie*	*Profitable truth-telling*	*Selfish lie*

Table 3-1. Definition of classes of advisor strategies

3.4.2. Taxonomy of mistrust and trust

Figure 3-4 displays the taxonomy of mistrust and trust for the investor based on their percentage extent of mistrust ($\overline{T}$ on the ordinate) and their percentage extent of suspected lying (L_{guess} on the abscissa). Here the dotted line represents strategies in which the investor expects to make an optimal investment

($F_{I;guess} = 0$). Hence, an investor with a strategy below this line would expect to underinvest ($F_{I;guess} < 0$), while an investor with a strategy above it would expect to make an overinvestment ($F_{I;guess} > 0$). Analogous to the advisor, I use this distinction (of different types of $F_{I;guess}$) in combination with the investor's (mis)trusting behavior ($\bar{T}$) to define classes of feasible strategies for the investor.[51]

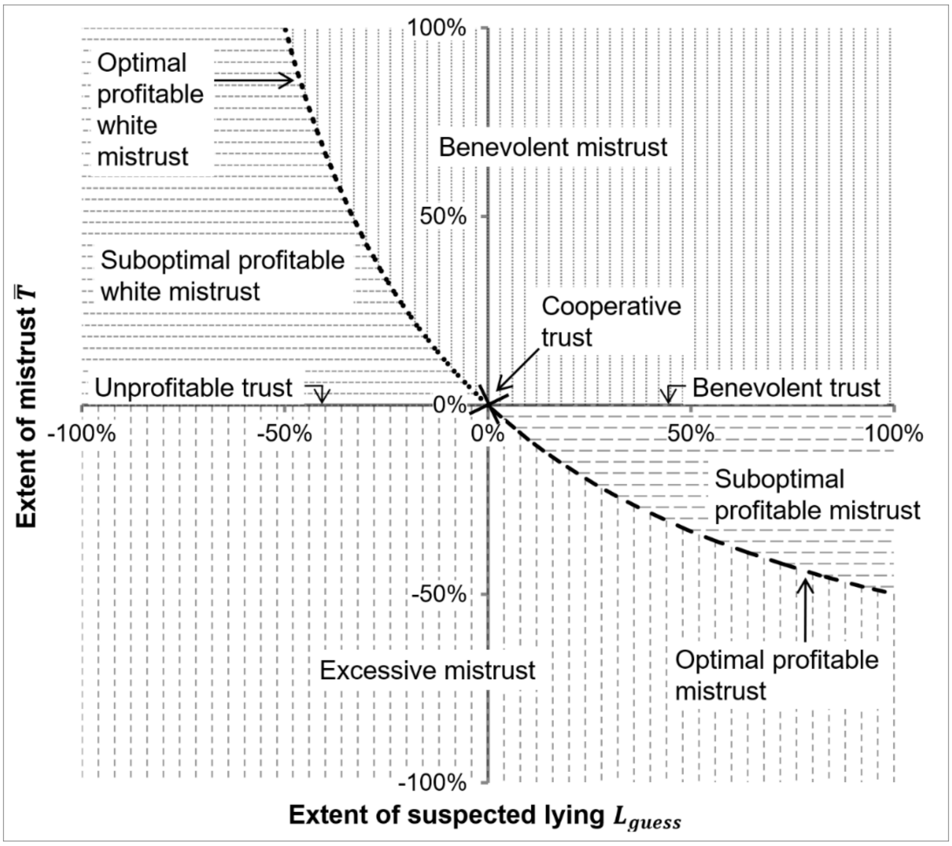


Figure 3-4. Taxonomy of mistrust and trust for the investor

As summarized in Table 3-2, there are the following nine classes of feasible strategies for the investor:

Definition 17. *Excessive mistrust*: The investor engages in risk-reducing mistrust ($\bar{T} < 0$) and therefore expects to make an underinvestment ($F_{I;guess} < 0$). Compared to more trusting behavior, the investor would expect this to reduce both players' payoffs.

Definition 18. *Optimal profitable mistrust*: The investor engages in risk-reducing mistrust ($\bar{T} < 0$) while suspecting the advisor to have overstated the true value of the optimal investment to an equal extent. Hence, the investor expects to make an optimal investment ($F_{I;guess} = 0$).

[51] It should be reminded that, in line with previous sender-receiver games (e.g., Peeters et al., 2013; Sutter, 2009), in the CDG "trust" refers to the investor's trusting *behavior* ($\bar{T}$) rather than their trusting *beliefs* (L_{guess}).

Definition 19. *Suboptimal profitable mistrust*: The investor engages in risk-reducing mistrust ($\bar{T} < 0$) while suspecting the advisor to have overstated the true value of the optimal investment to an even stronger extent. As a result, the investor expects to overinvest ($F_{I;guess} > 0$).

Definition 20. *Unprofitable trust*: The investor behaves completely trusting ($\bar{T} = 0$) even though he or she suspects the advisor to have understated the true value of the optimal investment. Thus, the investor expects to make an underinvestment ($F_{I;guess} < 0$).

Definition 21. *Cooperative trust*: The investor behaves completely trusting ($\bar{T} = 0$) believing that the advisor has told the truth. As a consequence, the investor expects to make an optimal investment ($F_{I;guess} = 0$).

Definition 22. *Benevolent trust*: The investor behaves completely trusting ($\bar{T} = 0$) even though he or she suspects the advisor to have overstated the true value of the optimal investment. Thus, the investor expects to overinvest ($F_{I;guess} > 0$).

Definition 23. *Suboptimal profitable white mistrust*: The investor engages in risk-seeking mistrust ($\bar{T} > 0$) while suspecting the advisor to have understated the true value of the optimal investment to an even stronger extent. As a result, the investor expects to make an underinvestment ($F_{I;guess} < 0$).

Definition 24. *Optimal profitable white mistrust*: The investor engages in risk-seeking mistrust ($\bar{T} > 0$) while suspecting the advisor to have understated the true value of the optimal investment to an equal extent. Hence, the investor expects to make an optimal investment ($F_{I;guess} = 0$).

Definition 25. *Benevolent mistrust*: The investor engages in risk-seeking mistrust ($\bar{T} > 0$) and therefore expects to make an overinvestment ($F_{I;guess} > 0$). Compared to more trusting behavior, the investor would expect this to reduce their payoff while increasing that of the advisor.

		(Mis)trusting behavior		
		Risk-reducing mistrust ($\bar{T} < 0$)	Trusting behavior ($\bar{T} = 0$)	Risk-seeking mistrust ($\bar{T} > 0$)
Expected investment	Underinvestment ($F_{I;guess} < 0$)	*Excessive mistrust*	*Unprofitable trust*	*Suboptimal profitable white mistrust*
	Optimal investment ($F_{I;guess} = 0$)	*Optimal profitable mistrust*	*Cooperative trust*	*Optimal profitable white mistrust*
	Overinvestment ($F_{I;guess} > 0$)	*Suboptimal profitable mistrust*	*Benevolent trust*	*Benevolent mistrust*

Table 3-2. Definition of classes of investor strategies

Some of the presented strategies appear more reasonable than others. For instance, why would one lie or behave mistrustingly if he or she expects their strategy to reduce both players' payoffs? Most certainly some strategies are more likely than others. With that in mind, in the next subsection, I define rational strategies from a game theoretical perspective.

3.5. Rational strategies

From a game theoretical perspective, some strategies are rational and others are not. Against this backdrop, appendix B.1. is dedicated to identifying rational strategies in the CDG. Therefore, in this appendix, I solve the CDG by finding its set of game theoretical equilibria, which allows me to determine strategies that are more likely to be pursued by rational players. Following this idea, I define *rational strategies* as game theoretical equilibrium strategies. My analysis is based on some basic assumptions: Firstly, both players are modeled as risk-neutral rational players who seek to maximize their expected utility based on their beliefs about the other player. Secondly, both players are assumed to value their monetary payoffs. Thirdly, I assume that the advisor has a preference for honesty, whereas the investor has a preference for trust.[52] Finally, and in line with basic game theoretical assumptions, I suppose that both players' beliefs about each other are correct in equilibrium.

Based on these assumptions, my analysis in appendix B.1. shows, on the one hand, that *rational advisors* would engage in either selfish lying, optimal profitable white lying, or cooperative truth-telling (with $F_{A;guess} \geq 0$ and $\bar{T}_{guess} \leq 0$).[53] However, which strategy they choose depends on their individual preference for honesty and their first-order beliefs about the investor's mistrust. More precisely, I find that the more (risk-reducing) mistrust a rational advisor expects from the investor ($|\bar{T}_{guess}| \uparrow$ with $\bar{T}_{guess} \leq 0$), the more he or she lies by overstating ($L \uparrow$ with $L \geq 0$).[54] On the other hand, my analysis reveals that *rational investors* would engage in either profitable mistrust, cooperative trust, or benevolent trust (with $F_{I;guess} \geq 0$ and $\bar{T} \leq 0$).[55] Here, their choice of strategy depends on their individual preference for trust and their beliefs about the advisor's honesty. In particular, I show that a higher extent of suspected lying ($L_{guess} \uparrow$ with $L_{guess} \geq 0$) makes a rational investor consider more mistrusting strategies (i.e., a higher possible extent of risk-reducing mistrust $|\bar{T}|$ with $\bar{T} \leq 0$).[56]

It can be concluded that rational players in the CDG are expected to base their lying and mistrusting behavior on their beliefs about the other player. This allows the identification of rational strategies for both players. These strategies can serve as a reference for the question of which feasible strategies are most likely to be pursued in the CDG.

---

[52] As discussed in the literature section of this paper, these assumptions are based on empirical work that suggests that people have a preference for honest behavior (e.g., Erat & Gneezy, 2012; Fischbacher & Föllmi-Heusi, 2013; Gneezy, 2005; Lundquist et al., 2009; Mazar et al., 2008; Vanberg, 2008) and a preference for trust (e.g., Barefoot et al., 1998; Gurtman, 1992; Kuroki, 2011; Sapienza et al., 2013). For more details, also refer to appendix B.1.2.
[53] For proof, see appendices B.1.2.1. to B.1.2.3.
[54] For proof, see appendix B.1.2.4.
[55] For proof, see appendices B.1.2.1. to B.1.2.3.
[56] For proof, see appendix B.1.2.4.

4. Hypotheses

An important aspect of this paper is to introduce the CDG and explore which strategies players pursue in this new experimental design. Due to the novelty of the game, this analysis is mostly explorative. However, in this section, I will briefly formulate my expectations towards both players' strategies in the CDG.

In the first place, based on my game theoretical analysis, I expect that most players will pursue mainly rational strategies. Therefore, I predict that the proportion of rational advisor/investor strategies among all of their pursued strategies will be significantly higher than its expected value based on random choices. On this basis, I formulate my first pair of hypotheses:

> **Hypothesis 1.** *The advisors pursue rational strategies (with $F_{A;guess} \geq 0$ and $\bar{T}_{guess} \leq 0$) disproportionately more often than other strategies (with $F_{A;guess} < 0$ or $\bar{T}_{guess} > 0$).*

> **Hypothesis 2.** *The investors pursue rational strategies (with $F_{I;guess} \geq 0$ and $\bar{T} \leq 0$) disproportionately more often than other strategies (with $F_{I;guess} < 0$ or $\bar{T} > 0$).*

In the second place, I expect that both players will engage in strategic decision making. Thus, under the assumption that strategic decision making when lying is reflected in the relationship between the displayed lying behavior and beliefs about others (e.g., López-Pérez & Spiegelman, 2013; Lundquist et al., 2009; Peeters et al., 2015), I conjecture that both players' behavior will be closely related to their first-order beliefs. This leads to my last two hypotheses:

> **Hypothesis 3.** *The percentage extent of expected risk-reducing mistrust ($\bar{T}_{guess}$ with inverted sign) is positively correlated with the percentage extent of lying (L).*

> **Hypothesis 4.** *The percentage extent of suspected lying (L_{guess}) is positively correlated with the percentage extent of risk-reducing mistrust ($\bar{T}$ with inverted sign).*

Before testing these hypotheses, I will explain the precise implementation of my experimental design and experimental procedures in the next two sections.

5. Design

I conducted ten consecutive rounds of the CDG. Before the game started, every player was randomly assigned to one of the two roles: advisor or investor. The role of a player never changed throughout the entire experiment. To prevent common learning effects and backward induction between rounds, I used a perfect stranger design. Therefore, in each round every advisor was assigned to a different investor. In particular, players were rotated in such a way that no one was matched with the same player twice. This was known to all players.

Before the game started, I defined its parameters. Firstly, to ensure that a change in the investment would equally affect both players' payoffs, I set both players' payoff factors (m_A and m_I) to 0.5.[57] Secondly, I introduced coins as an in-game currency. One hundred of these coins translated to 8 euros. Thirdly, based on that, I defined the overall maximal investment (i_{max}) to be equal to 100 coins. Lastly, I pre-generated the values of the optimal investments (i^*) for all ten rounds based on a random selection procedure.[58] The participants were informed about the random nature of the optimal investment values and it was pointed out to them that these values would most likely change between rounds. Each round's optimal investment was revealed to the advisors only at the beginning of the corresponding round. In order to prevent feedback-learning effects, the investors were informed about the values of all ten optimal investments only after the last round had been finished. Similarly, the advisors received the information about the values of their investors' investments only then, too. Finally, to determine each player's off-game payoff, one round was selected randomly at the end of the experiment. This procedure was introduced to the participants in advance. In my experiment the sixth round was selected to determine the off-game payoffs.

6. Experimental procedures

My experiment was conducted on February 1, 2018 at the University of Kassel with a total of 65 participants.[59] However, three participants did not finish the experiment and were therefore excluded ex post from the sample. Thus, 62 subjects are remaining – 25 females and 37 males with an average age of 21.89 years. In addition, the data obtained in my post-experimental questionnaire shows that 58.1% of my participants were students in economics, 17.7% in engineering, 9.7% in cultural studies, and 14.5% in other fields of study. Moreover, as my subjects were recruited from among participants of a basic course on game theory, they can be expected to have a basic understanding of strategic decision making in an economic setting. Even though this is not necessary for understanding or playing the CDG,

---

[57] By using equal payoff factors (m_A and m_I) the ratio between the expected profits from lying to the advisor (i.e., the sender) and the associated costs to the investor (i.e., the receiver) is equal to 1. This makes my results easier to compare to those of Gneezy (2005), since in most treatments of his sender-receiver game he implemented this same profit-loss ratio.

[58] To simplify the amounts of the optimal investments, I only allowed optimal investments that were divisible by five. The ten optimal investments, which I used in the ten rounds of my experiment, were: 50, 70, 25, 35, 10, 50, 70, 25, 35, and 10 coins in this order. Note that the first five optimal investment values were selected randomly. In the last five rounds, I reused the random optimal investment values of the previous five rounds. This was done in order to be able to analyze the temporal consistency of both players' behavior and first-order beliefs with identical information input. On this basis, in appendix B.4., I find that most players pursued similar strategies in two different rounds when they received identical information about the value of the optimal investment in these rounds. This shows that both players' decisions in the CDG are largely consistent over time.

[59] Since lying and mistrust are measured on continuous scales, this experiment requires a significantly lower sample size to provide reliable results than binary choice-based sender-receiver games. In fact, according to *GPower* (version 3.1.9.7), a sample size of 11 subjects *per group* would already achieve a statistical power of 0.805 to detect a significant difference in means between the advisors' extent of lying and the investors' extent of suspected lying at the 5%-threshold in one single round of the CDG. This is assuming that the extent of (suspected) lying is normally distributed and that the values of the population parameters are equal to the statistics of my sample. Under the same assumptions but with my actual sample size, the statistical power to detect a significant difference between the means of the average extents of lying and suspected lying at the 5%-threshold is 0.998.

I consider this an advantage, since this paper aims to investigate the lying and mistrusting behavior of people in an economic context that requires strategic thinking.

Since in a single round each player has to provide only two inputs, one round could easily be conducted with pen and paper. However, since the player rotation procedure through the ten rounds of my experiment was rather complex, I implemented it by using an online tool to conduct interactive experiments, *classEx* (Giamattei & Lambsdorff, 2019). This tool allows participants to log themselves into the experiment anonymously via their smartphones and make their decisions on screen while sitting in the lab.

My experimental procedure consisted of three stages: (1) the instruction stage, (2) the game stage, and (3) the post-experimental questionnaire. In (1), the instruction stage, every participant randomly received a sheet of paper with a unique ID that was used to assign them their role. Subsequently, I introduced *classEx* to all participants. After all participants had logged themselves into my experiment in *classEx* and had entered their ID, they read their instructions for the up-coming games on their screens. The participants were allowed to ask questions privately. In (2), the game stage, I conducted ten rounds of the CDG as described in the design section. All instructions and input screens of the CDG can be found in appendix B.2. The game stage took about 15 minutes, with each round taking less than 90 seconds to complete. After the last round was finished, each participant had to fill out (3), my post-experimental questionnaire. This was also done by using *classEx*. The active participation of the participants ended after they had completed the questionnaire. Up to this point, the experimental procedure took about 35 minutes. Note that the completion of the questionnaire was a necessary condition for receiving the full payoff at the end of the experiment.[60] The payoffs ranged from 0.8 to 10 euros with an average of 5.63 euros.

7. Results

In this section, I report my findings from the CDG. In the first place, I analyze the behavior and first-order beliefs of the players separately. Therefore, I use the seven key variables of the CDG (*7.1.*). In the second place, I analyze the relationship between both players' behavior and first-order beliefs (*7.2.*). This allows me to show which strategies the players' pursued in the CDG. Finally, I explore how the advisors took their potential to manipulate the investors into account when making their decisions (*7.3.*). The results section concludes with a short summary (*7.4.*).

7.1. Key variables

This subsection explores both players' behavior and first-order beliefs in the CDG by analyzing the seven key variables of the CDG. I begin by examining the advisors' extent of lying and the investors'

[60] If the participants did not complete their questionnaire, they received only half of their original payoff.

extent of suspected lying (*7.1.1.*). Then, I turn to the investors' extent of mistrust and the advisors' extent of expected mistrust (*7.1.2.*). Lastly, I analyze both players' extents of expected misinvestment and the real extent of misinvestment (*7.1.3.*). For this first analysis of the seven key variables, all ten rounds of the CDG are considered jointly. This is done by using averages of repeated measurements for each player. As the averages for individual players are statistically independent within player groups, this allows me to use statistical tests that assume independent observations in this section.

7.1.1. Extent of (suspected) lying

Table 3-3 compares the observed lying behavior of the advisors to the first-order beliefs of the investors towards it. Firstly, the table shows the proportions of different types of (suspected) lies on average over all ten rounds.[61] Secondly, it displays the average percentage extent of (suspected) lying over all ten rounds.

Concerned variables	Lying (advisor)	Suspected lying (investor)	Difference of averages (1st-2nd)	p-value (two-sided Mann-Whitney U test[1])
Proportion of (suspected) lying	78.39%	78.39%	0%	0.960
...by understating	*1.29%*	*20.00%*	*-18.71%*	*< 0.001*
...by overstating	*77.10%*	*58.39%*	*18.71%*	*0.004*
Proportion of (suspected) truthful advice	21.61%	21.61%	0%	0.960
Extent of (suspected) lying	148.10%	55.94%	92.16%	< 0.001

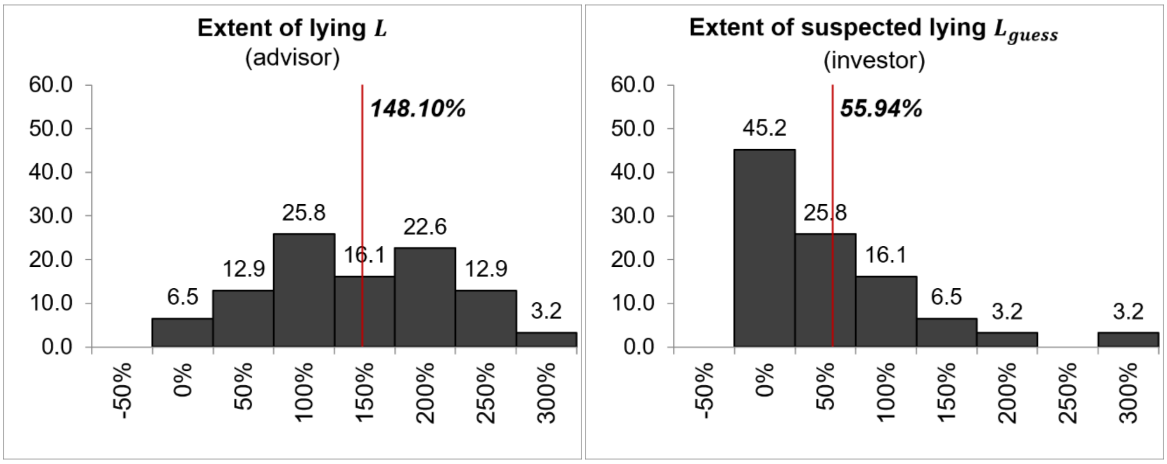
[1] Mann-Whitney U test to compare the mean rank of the respective concerned variables between both players.

Table 3-3. Proportion and extent of (suspected) lying

As can be seen, exactly as many pieces of advice were lies (78.39% of all advice) as there were suspected to be by the investors (78.39% of all advice). As a result, the proportions of actual and suspected lies do not differ significantly (Mann-Whitney U test: p = 0.960). Solely based on this information, one might (falsely) conclude that the investors' beliefs about their advisors' dishonesty were highly accurate. However, taking the extents of lying and suspected lying into account reveals that this assessment is not correct.[62] To show why, Figure 3-5 visualizes the distributions of the percentage extent of lying (L) on the left (3-5a) and of suspected lying (L_{guess}) on the right (3-5b).

[61] The *proportion of lying* refers to the percentage of advisors who lied (i.e., the relative frequency of observed $L \neq 0$), calculated as an average over all ten rounds. Analogously, the *proportion of suspected lying* refers to the percentage of investors who suspected their advisors to have lied (i.e., the relative frequency of observed $L_{guess} \neq 0$), calculated as an average over all ten rounds.

[62] This is important since the findings of binary choice-based sender-receiver games (e.g., Gneezy, 2005; López-Pérez & Spiegelman, 2013; Peeters et al., 2015; Sánchez-Pagés & Vorsatz, 2007) are usually based solely on the proportions of liars and truth-tellers.



(a) **(b)**

Figure 3-5. Distributions of the extent of lying and the extent of suspected lying: (a) Extent of lying L; (b) Extent of suspected lying L_{guess}

The comparison of both distributions shows that the advisors' percentage extent of lying differs from the investors' percentage extent of suspected lying on various levels: In the first place, both samples do not follow the same distribution (two-sample Kolmogorov-Smirnov test: $p < 0.001$). While the distribution of the percentage extent of lying appears to be close to a normal distribution (one-sample Kolmogorov-Smirnov test: $p = 0.200$), the distribution of the percentage extent of suspected lying does not (one-sample Kolmogorov-Smirnov test: $p = 0.009$). In fact, the latter is decreasing almost monotonously. The evident difference between both distributions indicates that the investors followed an incorrect pattern to estimate their advisors' lying behavior. In the second place, the peak of the frequency distribution of the investors' percentage extent of suspected lying is at the distribution's lower limit around zero (at around $L_{guess} = 0$). By contrast, the frequency distribution of the advisors' percentage extent of lying peaks twice at high levels ($L \gg 0$), once around 100% and once around 200%. This suggests that truthful behavior was a much weaker reference point for the advisors than the investors expected it to be. Finally, the average percentage extent to which the advisors lied (148.10%) is significantly higher than the average percentage extent to which the investors suspected them to do so (55.94%) (Mann-Whitney U test: $p < 0.001$). It follows that the investors underestimated the percentage extent of lying on average by 92.16% (L vs. L_{guess}).

> **Finding 1.** *While the investors correctly predicted the proportion of liars, they largely underestimated the extent of lying ($L > L_{guess}$).*

7.1.2. Extent of (expected) mistrust

Table 3-4 contrasts the expectations that the advisors had about their investors' mistrust with the actual mistrust of the investors. To begin with, it displays the proportions of different types of (expected) mistrust on average over all ten rounds.[63] In addition, the table reports the average percentage extent of (expected) mistrust over all ten rounds.

Concerned variables	Expected mistrust (advisor)	Mistrust (investor)	Difference of averages (2nd-1st)	p-value (two-sided Mann-Whitney U test[1])
Proportion of (expected) mistrust	68.71%	72.26%	3.55%	0.701
Proportion of (expected) risk-reducing mistrust	*58.06%*	*57.10%*	*-0.96%*	*1.000*
Proportion of (expected) risk-seeking mistrust	*10.65%*	*15.16%*	*4.51%*	*0.273*
Proportion of (expected) trusting investments	31.29%	27.74%	-3.55%	0.701
Extent of (expected) mistrust[2]	-8.01%	-7.21%	0.80%	0.767

[1] Mann-Whitney U test to compare the mean rank of the respective concerned variables between both players; [2] Note that a negative percentage extent of (expected) mistrust refers to (expectations of) risk-reducing mistrust.

Table 3-4. Proportion and extent of (expected) mistrust

It can be seen that the advisors expected 68.71% of all investments to be mistrusting, while 72.26% of them actually were. This minor difference is not significant (Mann-Whitney U test: p = 0.701), which suggests that the advisors predicted the proportion of mistrust highly accurately. For a more detailed analysis, Figure 3-6 illustrates the distributions of the advisors' percentage extent of expected mistrust ($\bar{T}_{guess}$) on the left (3-6a) and the investors' percentage extent of mistrust ($\bar{T}$) on the right (3-6b). To read this figure, recall that a negative percentage extent of (expected) mistrust refers to (expectations of) risk-reducing mistrust, whereas a positive percentage refers to (expectations of) risk-seeking mistrust.

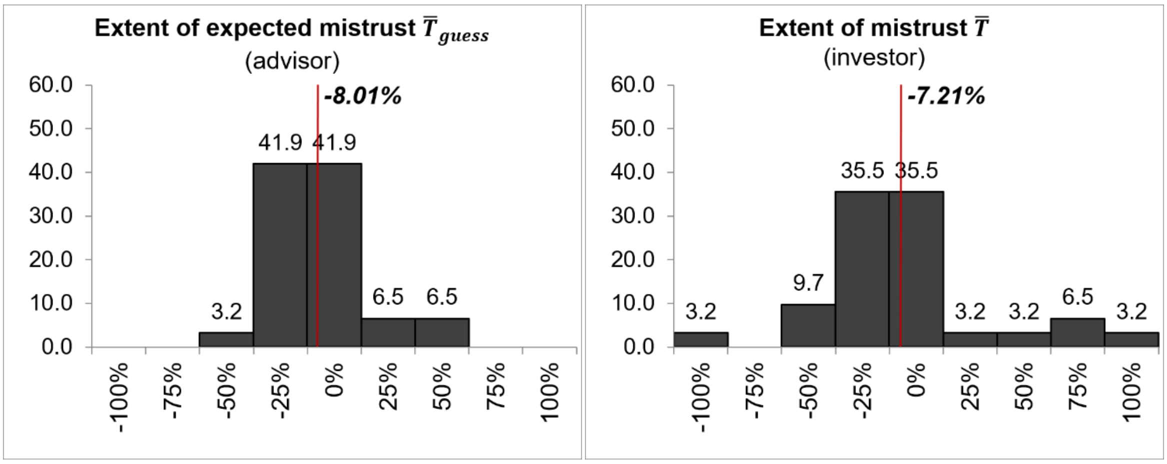


Figure 3-6. Distributions of the extent of expected mistrust and the extent of mistrust: (a) Extent of expected mistrust $\bar{T}_{guess}$; (b) Extent of mistrust $\bar{T}$

As can be seen by comparing both distributions, the advisors estimated their investors' percentage extent of mistrust highly accurately: Firstly, both samples appear to come from a similar distribution (a two-

[63] The *proportion of mistrust* refers to the percentage of investors who did not follow their received advice (i.e., the relative frequency of observed $\bar{T} \neq 0$), calculated as an average over all ten rounds. Moreover, the *proportion of expected mistrust* refers to the percentage of advisors who expected their investors to make mistrusting investments (i.e., the relative frequency of observed $\bar{T}_{guess} \neq 0$), calculated as an average over all ten rounds.

sample Kolmogorov-Smirnov test is not significant: p = 0.607). Secondly, both frequency distributions peak close to zero (at around $\bar{T}_{guess} = 0$ and $\bar{T} = 0$), which indicates that trusting behavior was as much of a reference point for the investors as the advisors expected. Thirdly, the average percentage extent of expected mistrust (-8.01%) barely differs from the average percentage extent of actual mistrust (-7.21%). This extraordinarily small difference ($\bar{T}_{guess}$ vs. $\bar{T}$) amounts to only 0.80% and is not significant (Mann-Whitney U test: p = 0.767).

Finding 2. *The advisors predicted both the proportion and the extent of mistrust highly accurately ($\bar{T}_{guess} \approx \bar{T}$).*

7.1.3. Extent of (expected) misinvestment

Table 3-5 compares the expectations of the advisors and the investors about the overall quality of investments. Firstly, it displays the proportions of different types of expected misinvestments for both players on average over all ten rounds.[64] Secondly, the table shows both players' average percentage extent of expected misinvestment over all ten rounds.

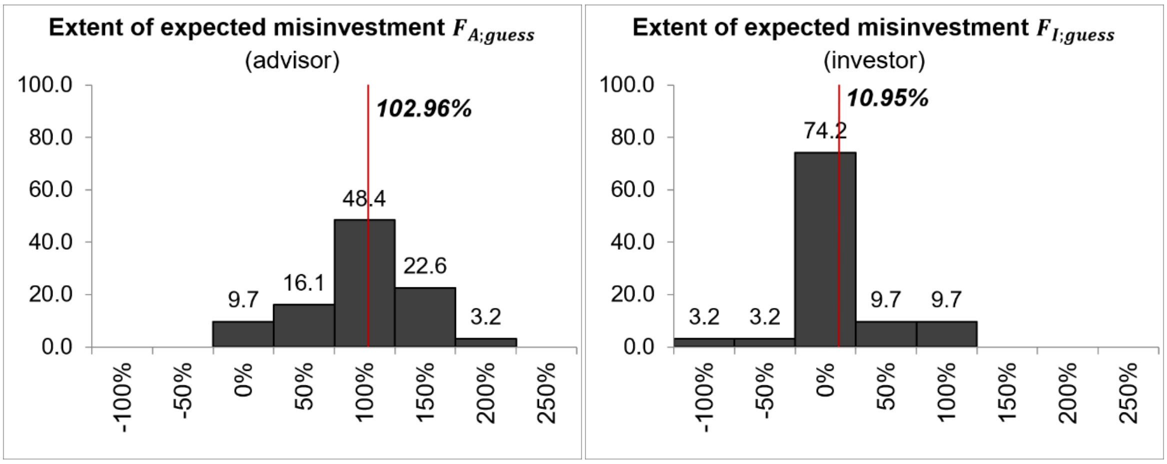
Concerned variables	Expected misinvestment (advisor)	Expected misinvestment (investor)	Difference of averages (1st-2nd)	p-value (two-sided Mann-Whitney U test[1])
Proportion of expected misinvestments	73.22%	57.42%	15.80%	0.086
Proportion of expected underinvestments	*11.61%*	*28.39%*	*-16.78%*	*0.008*
Proportion of expected overinvestments	*61.61%*	*29.03%*	*32.58%*	*<0.001*
Proportion of expected optimal investments	26.78%	42.58%	-15.80%	0.086
Extent of expected misinvestment	102.96%	10.95%	92.01%	<0.001

[1] Mann-Whitney U test to compare the mean rank of the respective concerned variables between both players.

Table 3-5. Proportion and extent of expected misinvestment

The advisors expected a moderately and non-significantly higher proportion of non-optimal investments than the investors (73.22% vs. 57.42% of all investments; Mann-Whitney U test: p = 0.086). However, both players expected different types of misinvestments: Whereas the advisors expected their investors to make mostly overinvestments, the investors expected an approximately equal number of over- and underinvestments. As a consequence, the advisors expected a significantly lower proportion of under-investments (11.61% vs. 28.39% of all investments; Mann-Whitney U test: p = 0.008) and a significantly higher proportion of overinvestments than the investors (61.61% vs. 29.03% of all investments; p < 0.001). Going into more detail, Figure 3-7 visualizes the distributions of the percentage extent of expected misinvestment of the advisors ($F_{A;guess}$) on the left (3-7a) and of the investors ($F_{I;guess}$) on the right (3-7b).

[64] The *proportion of expected misinvestments* refers to the percentage of advisors (or investors) who expected the investments to be non-optimal (i.e., the relative frequency of observed $F_{A;guess} \neq 0$ or $F_{I;guess} \neq 0$, respectively), calculated as an average over all ten rounds.



(a) **(b)**

Figure 3-7. Distributions of both players' extent of expected misinvestment: (a) Extent of expected misinvestment of the advisor $F_{A;guess}$; (b) Extent of expected misinvestment of the investor $F_{I;guess}$

Comparing both distributions shows that the average percentage extent to which the advisors expected their investors to overinvest (102.96%) is significantly higher than the average percentage extent to which the investors expected to do so (10.95%) (Mann-Whitney U test: $p < 0.001$). It follows that the advisors expected the average percentage extent of misinvestment to be 92.01% higher than the investors ($F_{A;guess}$ vs. $F_{I;guess}$).

Finding 3. *The advisors expected more overinvestments and a larger extent of misinvestment than the investors ($F_{A;guess} > F_{I;guess}$).*

Concerned variables	Value
Proportion of misinvestments	83.87%
Proportion of underinvestments	*24.52%*
Proportion of overinvestments	*59.35%*
Proportion of optimal investments	16.13%
Extent of misinvestment	88.96%

Table 3-6. Proportion and extent of misinvestment

In order to assess the quality of both players' estimates of the outcomes of the investments, Table 3-6 provides an overview of the actual quality of investments. In the first place, it shows the proportions of different types of investments on average over all ten rounds.[65] In the second place, it displays the average percentage extent of misinvestment over all ten rounds. As can be seen, only 16.13% of all investments were optimal. This is because 24.52% of all investments were underinvestments and 59.35% overinvestments.

[65] The *proportion of misinvestments* refers to the percentage of investors who made non-optimal investments (i.e., the relative frequency of observed $F \neq 0$), calculated as an average over all ten rounds.

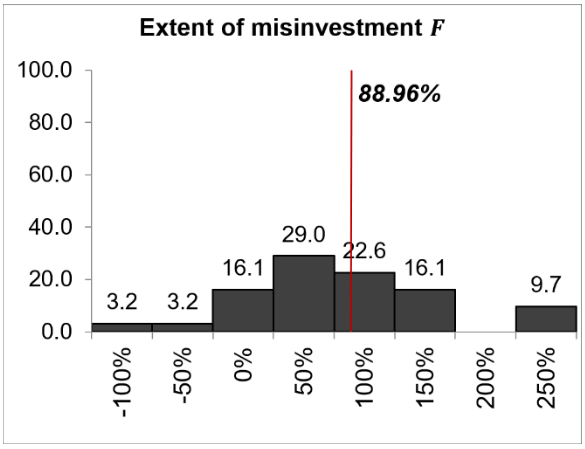


Figure 3-8. Distribution of the extent of misinvestment F

For a more detailed analysis, Figure 3-8 visualizes the distribution of the percentage extent of misinvestment (F). The comparison of the observed quality of investments and both players' expectations towards it reveals that the advisors' estimates of the extent of misinvestment were much more accurate than those of the investors: On the one hand, on average the advisors expected only a moderately and non-significantly higher percentage extent of misinvestment than there was ($F_{A;guess}$ vs. F; 102.96% vs. 88.96%; Mann-Whitney U test: p = 0.229). As a result, they overestimated the percentage extent of misinvestment on average by only 14.00%. On the other hand, the investors significantly underestimated the percentage extent of misinvestment by 78.01% on average ($F_{I;guess}$ vs. F; 10.95% vs. 88.96%; Mann-Whitney U test: p < 0.001).

> **Finding 4.** *The advisors barely overestimated the extent of misinvestment ($F_{A;guess} \approx F$), while the investors largely underestimated it ($F_{I;guess} < F$).*

7.2. Strategy analysis: the relationship between lying, mistrust, and first-order beliefs

In this subsection, I examine which strategies both players pursued in the CDG. Firstly, I analyze the relationship between both players' behavior and their first-order beliefs. Secondly, I examine which types of strategies both players chose. I will begin with the advisors (*7.2.1.*) and then turn to the investors (*7.2.2.*). For this analysis, observations from repeated measurements are considered individually. Note here that repeated observations of the same players in different rounds of the CDG are *not* statistically independent. To deal with this, all tests on data with observations from different rounds are based on bootstrap procedures that account for within correlations between individual decisions in different rounds. More details on these bootstrap procedures can be found in appendix B.5.

7.2.1. Lying and expected mistrust (advisors)

Figure 3-9 visualizes the relationship between the advisors' lying behavior and their first-order beliefs about their investors.[66] It illustrates that the percentage extent of lying (L) significantly increases with the percentage extent of expected risk-reducing mistrust (negative $\bar{T}_{guess}$) (Spearman's rank correlation[67] between L and $-\bar{T}_{guess}$: ρ = 0.397 with p < 0.001). This relationship provides support for *hypothesis H3*. In line with this, lying by overstating ($L > 0$) was observed significantly more often for advisors who expected risk-reducing mistrust ($\bar{T}_{guess} < 0$) than for advisors with other expectations (in 92.22% vs. 56.15% of cases; bootstrap unconditional proportion test[68]: p < 0.001). Moreover, advisors who expected to be trusted told the truth significantly more often than advisors who expected to be mistrusted (in 44.33% vs. 11.27% of cases; bootstrap unconditional proportion test: p < 0.001). The reported differences underline that the advisors' first-order beliefs and their lying behavior are closely related to one another. This indicates that the advisors tended to make rather strategic decisions.

> **Finding 5.** *The advisors' lying behavior (L) is closely related to their expectations of being mistrusted ($\bar{T}_{guess}$)* (supporting hypothesis H3).

[66] For the purpose of illustration, only the most relevant area of the plot in Figure 3-9 is displayed. Also note that the dotted line in the figure marks the hypothetical line on which the advisors expected the investments to be optimal (which would imply: $F_{A;guess} = 0$). While points below this line represent *expectations of under-investments* ($F_{A;guess} < 0$), points above it represent *expectations of overinvestments* ($F_{A;guess} > 0$).

[67] Note that, in this paper, the p-values of all correlations that are based on observations from different rounds are calculated using a bootstrap correlation coefficient test that accounts for within correlations between individual decisions in different rounds. For more details, see appendix B.5.

[68] For details on this bootstrap procedure, see appendix B.5.

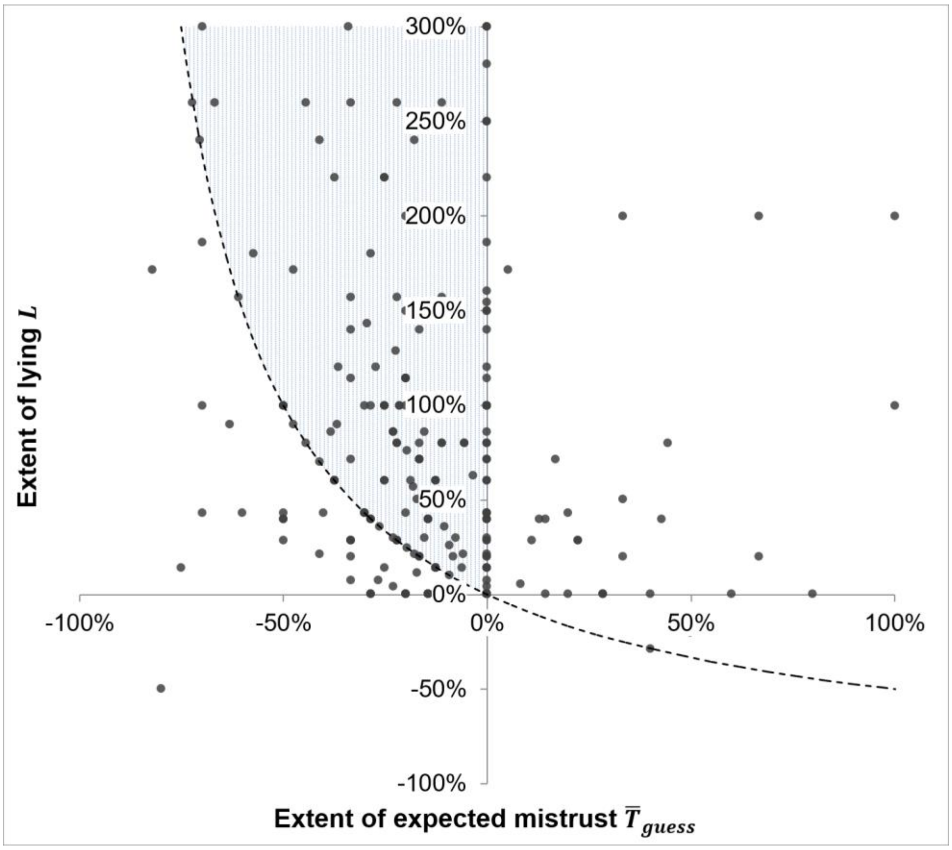


Figure 3-9. Relationship between the extent of lying L and the extent of expected mistrust $\bar{T}_{guess}$

To show which strategies the advisors pursued, Table 3-7 summarizes the proportions of *classes of observed advisor strategies* (which are also displayed in a non-aggregated form in Figure 3-9). This summary is based on the taxonomy of lies and truth-telling defined above.

		Lying behavior		
		Understating lie ($L < 0$)	Truth-telling ($L = 0$)	Overstating lie ($L > 0$)
Expected investment	Underinvestment ($F_{A;guess} < 0$)	*Spiteful lie* 0.32%	*Unprofitable truth-telling* 4.19%	*Suboptimal profitable white lie* 7.10%
	Optimal investment ($F_{A;guess} = 0$)	*Optimal altruistic white lie* 0.65%	*Cooperative truth-telling* 13.87%	*Optimal profitable white lie* 12.26%
	Overinvestment ($F_{A;guess} > 0$)	*Suboptimal altruistic white lie* 0.32%	*Profitable truth-telling* 3.55%	*Selfish lie* 57.74%
	Total	Understating lie 1.29%	Truth-telling 21.61%	Overstating lie 77.10%

Table 3-7. Proportions of advisor strategy classes

It can be seen that in most cases (57.74%) the advisors were *selfish liars*, i.e., liars who lied in order to make their investors overinvest, which would increase their own payoffs while reducing the payoffs of their investors. However, in 19.36% of cases the advisors told *profitable white lies*. This means that they lied by overstating the optimal investment in the expectation of preventing their investors (at least partially) from underinvesting. Thereby, they would increase both players' payoffs. Only in 0.97% of cases the advisors engaged in *altruistic white lying*, i.e., they lied by understating the optimal investment in order to prevent their investors (at least partially) from overinvesting. This strategy would reduce their own payoffs while increasing the payoffs of their investors. In even fewer cases (0.32%) the advisors told *spiteful lies*, i.e., lies that understated the optimal investment in order to make the investor underinvest, which would reduce both players' payoffs. In all other cases (21.61%) the advisors told the truth. Most of them were *cooperative truth-tellers* (in 13.87% of all cases). These advisors expected their investors to trust them and gave honest advice, which in turn would result in optimal investments. However, some advisors told the truth, even though they then expected non-optimal investments. In particular, in 4.19% of cases the advisors were *unprofitable truth-tellers*, i.e., advisors who were willing to accept underinvestments and, therefore, a reduction of both players' payoffs in order to tell the truth. By contrast, in 3.55% of cases the advisors were *profitable truth-tellers*, implying that they gave truthful advice and still expected their investors to overinvest. This would increase their own payoffs but reduce the payoffs of their investors.

Assuming that the advisors' behavior was consistent with their beliefs and their individual preferences for honesty, they pursued *rational strategies* in 77.74% of all cases. In Figure 3-9 this refers to all strategy points that are located within the hatched area, i.e., all points that are simultaneously on or above the dotted line ($F_{A;guess} \geq 0$) and on or left from the ordinate ($\bar{T}_{guess} \leq 0$). That includes all cases in which the advisors engaged in cooperative truth-telling or optimal profitable white lying as well as a major fraction of cases in which they told selfish lies.[69] It comes as no surprise that the advisors pursued these rational strategies more often than other strategies in every round (two-sided binomial tests: $p < 0.001$ for each round[70]). This is consistent with *hypothesis H1* and therefore supports the idea of rational decision making based on individual lie aversion and rational beliefs.

Finding 6. *The advisors pursued rational strategies disproportionately more often (in 77.74% of cases) than other strategies. In most of all cases (57.74%) the advisors were selfish liars* (supporting hypothesis H1).

[69] Note that 89.39% of all selfish liars pursued a rational strategy. The rest of them however had beliefs about their investors' behavior that are not rational from a game theoretical point of view.

[70] The probability $p_{A;rat}$ that an advisor would engage in a rational strategy, i.e., a potential equilibrium strategy (with $F_{A;guess} \geq 0$ and $\bar{T}_{guess} \leq 0$), by making random choices depends on the given values of the optimal (i^*) and the maximal (i_{max}) investment. If the maximal investment is given, this probability can be described by the following function of the optimal investment: $p_{A;rat}(i^*) = \frac{1}{i_{max}(i_{max}-i^*)} * \left[0.5 * {i_{max}}^2 + 0.5 * i^{*2} - i_{max} * i^*\right]$. With i_{max} = 100, the values of $p_{A;rat}(i^*)$ for the five optimal investments i^*, which I used in my experiment, are: $p_{A;rat}(50) = 0.125$, $p_{A;rat}(70) = 0.045$, $p_{A;rat}(25) = 0.281$, $p_{A;rat}(35) = 0.211$, and $p_{A;rat}(10) = 0.405$. Based on these probabilities, I conducted a separate two-sided binomial test for each round to compare the proportion of rational strategies to its expected value based on random choices.

7.2.2. Mistrust and suspected lying (investors)

The scatter plot in Figure 3-10 shows the relationship between the investors' mistrusting behavior and their first-order beliefs about their advisors.[71] It can be seen that the percentage extent of risk-reducing mistrust (negative $\bar{T}$) significantly increases with the percentage extent of suspected lying (L_{guess}) (Spearman's rank correlation between $-\bar{T}$ and L_{guess}: $\rho = 0.752$ with $p < 0.001$). This is consistent with *hypothesis H4*. In addition, investors who suspected their advisors to have lied by overstating the value of the optimal investment ($L_{guess} > 0$) engaged significantly more often in risk-reducing mistrust ($\bar{T} < 0$) than investors with other expectations (in 87.85% vs. 13.95% of cases; bootstrap unconditional proportion test: $p < 0.001$). In line with this, investors who suspected that their advisors told the truth engaged in completely trusting behavior significantly more often than investors who suspected to be lied to (in 79.10% vs. 13.58% of cases; bootstrap unconditional proportion test: $p < 0.001$). These large differences highlight the fact that the investors' first-order beliefs and their mistrusting behavior are closely related to one another, which suggests that the investors engaged in rather strategic decision making.

> **Finding 7.** *The investors' mistrusting behavior ($\bar{T}$) is closely related to their expectations of being lied to by their advisors (L_{guess})* (supporting hypothesis H4).

---

[71] For the purpose of illustration, only the most relevant area of the plot in Figure 3-10 is displayed. Also note that the dotted line in the figure marks the hypothetical line on which the investors expected to make optimal investments (which would imply: $F_{I;guess} = 0$). Hence, points below this line represent *expectations of under-investments* ($F_{I;guess} < 0$), whereas points above it represent *expectations of overinvestments* ($F_{I;guess} > 0$).

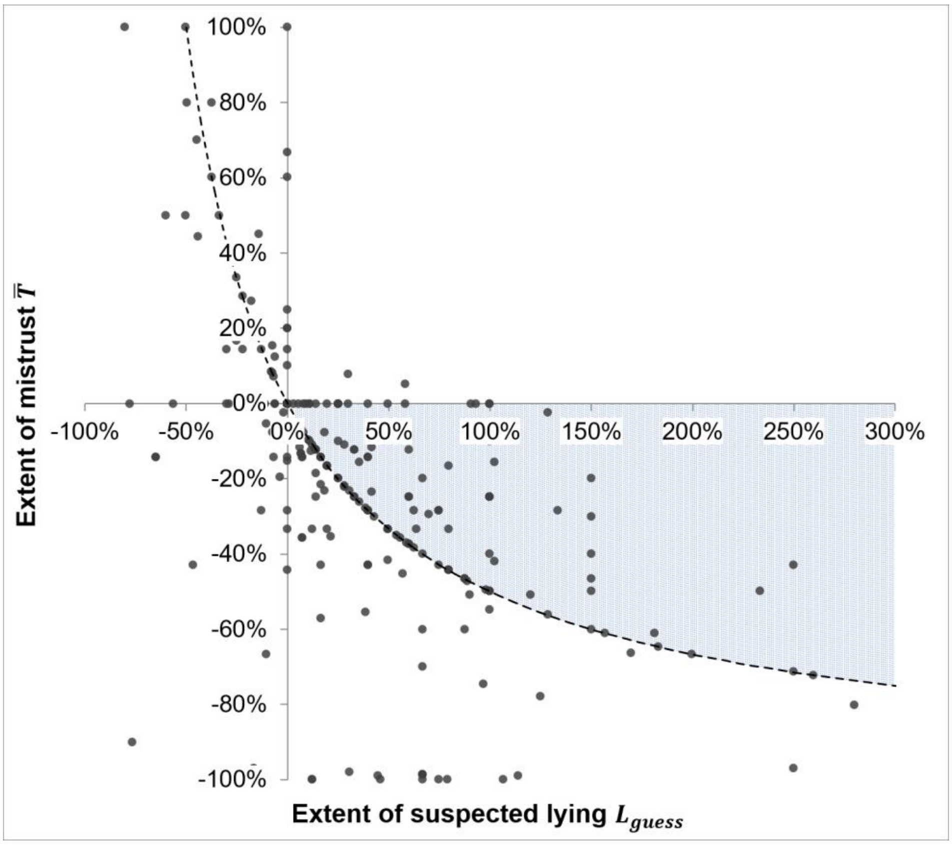


Figure 3-10. Relationship between the extent of mistrust $\bar{T}$ and the extent of suspected lying L_{guess}

To illustrate which strategies the investors chose, Table 3-8 provides an overview of the proportions of *classes of observed investor strategies* (which are also displayed in a non-aggregated form in Figure 3-10). This summary is based on the earlier introduced taxonomy of mistrust and trust.

		(Mis)trusting behavior		
		Risk-reducing mistrust ($\bar{T} < 0$)	Trusting behavior ($\bar{T} = 0$)	Risk-seeking mistrust ($\bar{T} > 0$)
Expected investment	Underinvestment ($F_{I;guess} < 0$)	*Excessive mistrust* *20.65%*	*Unprofitable trust* *4.19%*	*Suboptimal profitable white mistrust* *3.55%*
Expected investment	Optimal investment ($F_{I;guess} = 0$)	*Optimal profitable mistrust* *20.65%*	*Cooperative trust* *17.10%*	*Optimal profitable white mistrust* *4.84%*
Expected investment	Overinvestment ($F_{I;guess} > 0$)	*Suboptimal profitable mistrust* *15.80%*	*Benevolent trust* *6.45%*	*Benevolent mistrust* *6.77%*
	Total	Risk-reducing mistrust *57.10%*	Trusting behavior *27.74%*	Risk-seeking mistrust *15.16%*

Table 3-8. Proportions of investor strategy classes

As shown before, in most of the cases (57.10%) the investors engaged in risk-reducing mistrust. Most of them engaged in *profitable mistrust* (in 36.45% of all cases), which means that they suspected their advisors to have overstated the optimal investment and engaged in risk-reducing mistrust in order to (at least partially) improve the quality of their investments.[72] However, in 20.65% of cases the investors even engaged in risk-reducing mistrust in the expectation of making underinvestments. These investors expected their *excessive mistrust* to reduce both players' payoffs when compared to more trusting behavior. This indicates that, to some extent, they valued risk aversion over monetary gain. By contrast, in 8.39% of cases the investors engaged in *profitable white mistrust*, i.e., they suspected their advisors to have understated the optimal investment and engaged in risk-seeking mistrust in order to (at least partially) improve the quality of their investments, which would increase both players' payoffs. Moreover, in 6.77% of cases the investors engaged in *benevolent mistrust*, implying that they engaged in risk-seeking mistrust, even though they thereby expected to overinvest. Compared to more trusting behavior, these investors expected their behavior to reduce their own payoffs while increasing the payoffs of their advisors. In all the other cases (27.74%) the investors trusted their advisors. Most of them engaged in *cooperative trust* (in 17.10% of all cases). This means that they expected their advisors to have given truthful advice and exactly followed it, which would result in optimal investments. However, some investors followed their received advice, even though they then expected to make non-optimal investments. In particular, in 6.45% of cases the investors engaged in *benevolent trust*, i.e., they decided to trust their advisors and therefore expected to overinvest, which would reduce their own payoffs while increasing the payoffs of their advisors. By contrast, in 4.19% of cases the investors engaged in *unprofitable trust*, implying that they were willing to make an underinvestment and, therefore, to accept a reduction of both players' payoffs in order to behave completely trusting.

Under the assumption that the investors' behavior was consistent with their beliefs and their individual preferences for trust, they pursued *rational strategies* in 60.00% of all cases. In Figure 3-10 this refers to all strategy points that are located within the hatched area, i.e., all points that are simultaneously on or right from the dotted line ($F_{I;guess} \geq 0$) and on or below the abscissa ($\bar{T} \leq 0$). This includes all cases in which the investors engaged in cooperative trust, benevolent trust, or profitable mistrust. The investors pursued these rational strategies more often than other strategies (two-sided bootstrap binomial test[73]: p = 0.045). This finding is consistent with *hypothesis H2*, which speaks in favor of the conjecture

[72] Notably, in 25.49% of all cases the investors suspected their advisors to have lied but still expected to make optimal investments by perfectly compensating their advisors' extent of lying. In Figure 3-10, this refers to all strategy points that are located on the dotted line except for those at the point of origin.

[73] The probability $p_{I;rat}$ that an investor would engage in a rational strategy, i.e., a potential equilibrium strategy (with $F_{I;guess} \geq 0$ and $\bar{T} \leq 0$), by making random choices depends on the values of the received advice (a) and the maximal investment (i_{max}). If the maximal investment is given, this probability can be described by the following function of the received advice number: $p_{I;rat}(a) = 0.5 * \left(\frac{a}{i_{max}}\right)^2$.
With $a \in [0, i_{max}]$, this function maximizes for $a = i_{max}$. Given that $i_{max} = 100$, this yields: $p_{I;rat}(a) \leq p_{I;rat}(a = i_{max} = 100) = 0.5$. Since not all investors received the same advice, I use this upper limit of $p_{I;rat}$ to perform the most conservative two-sided bootstrap binomial test possible that compares the proportion of rational strategies to its expected value based on random choices. For more details on this bootstrap procedure, see appendix B.5.

that the investors engaged in rational decision making based on individual trust preferences and rational beliefs.

Finding 8. *The investors pursued rational strategies disproportionately more often (in 60.00% of cases) than other strategies. Most of them engaged in profitable mistrust* (supporting hypothesis H2).

7.3. Potential to manipulate others when lying

As reported in the previous subsection, most advisors told selfish lies with the aim of manipulating their investors into overinvesting. In addition, my results suggest that the advisors' behavior tended to be strategic. But how well thought out were the sizes of their lies? In this subsection, I explore how the advisors took into account their potential to manipulate their investors (*7.3.1.*) and whether their considerations were correct (*7.3.2.*).

7.3.1. Strategic deception (advisors)

In order to show how the advisors considered their potential to manipulate others, I first want to point out that they lied to the fullest possible extent in only 8.71% of cases. To explain this, it makes sense to analyze advisors who gave particularly high advice separately in more detail. Therefore, very high advice is assumed to be equal or superior to 85 coins. When considering only cases in which the advisors did *not* give such very high advice, the given advice is significantly and strongly correlated with the expected investments (Spearman's rank correlation: $\rho = 0.669$ with $p < 0.001$). This suggests that the advisors expected their investors to use advice below the defined threshold as a reference for the investments. However, this changes when only cases in which the advisors gave very high advice are considered. Here, the correlation between the given advice and expected investments is close to zero and not significant (Spearman's rank correlation: $\rho = -0.024$ with $p = 0.807$). The difference between these two correlations is significant (two-sample bootstrap correlation coefficient test[74]: $p < 0.001$). This indicates that the advisors expected their investors to particularly mistrust very high advice in such a way that giving higher advice above a certain level would not lead to higher investments and, therefore, not deliver a higher payoff. On this basis, it is reasonable to assume that the advisors took their potential to manipulate the investors into account when making their decisions, which in turn could explain why so many advisors refrained from lying to the fullest possible extent.

Finding 9. *The advisors very seldom lied to the fullest possible extent, as they expected their investors to particularly mistrust very high advice.*

[74] Just like the other bootstrap tests in this paper, this bootstrap procedure accounts for within correlations between individual observations in different rounds, which originate from repeated measurements. For more details, see appendix B.5.

7.3.2. *Predictable mistrust (investors)*

Most interestingly, the considerations of the advisors about their potential to manipulate the investors turned out to be correct: As expected by the advisors, the investors particularly mistrusted very high advice (i.e., advice equal or superior to 85). This is reflected in the fact that the correlation between the investments and the received advice is strong and highly significant when only cases in which the investors did *not* receive such very high advice are considered (Spearman's rank correlation: $\rho = 0.549$ with $p < 0.001$). By contrast, the same correlation is close to zero and not significant when considering only cases in which the received advice was very high (Spearman's rank correlation: $\rho = 0.040$ with $p = 0.731$). The difference between these two correlations is significant (two-sample bootstrap correlation coefficient test: $p < 0.001$). This suggests that the investors mistrusted particularly high advice by not following it any further above a certain level.

> **Finding 10.** *The investors mistrusted very high advice in such a way that they did not make higher investments when receiving higher advice above a certain level.*

It can be concluded that both players made strategic considerations based on the size of the lie and the size of mistrust.

7.4. Summary

The analysis of lying and mistrust in the CDG has shown that both players tended to make strategic and belief-based decisions. In particular, I find that their behavior was consistent with their first-order beliefs (*see Findings 5 and 7; Hypotheses 3 and 4*) and that most of them pursued rational strategies (*see Findings 6 and 8; Hypotheses 1 and 2*). As a consequence, both players' financial success in the game was largely determined by the accuracy of their first-order beliefs (*see Findings 3 and 4*). Here, the advisors took advantage of the fact that their first-order beliefs were much more accurate than those of the investors (*see Findings 1 and 2*). In fact, the advisors even correctly predicted that the investors would disproportionally mistrust very high advice (*see Findings 9 and 10*). As a result, most advisors avoided lying to the fullest possible extent. These findings suggest that the advisors took their potential to manipulate the investors into consideration when making their decisions. By this means, most of them successfully tricked their investors into overinvesting. In short, most advisors engaged in strategic deception while exploiting the predictability of their investors' mistrust.

These findings show that the size of the lie and the size of mistrust play an important role in strategic deception. Obviously, this is based on the assumption that both players' behavior can be interpreted on continuous scales. To test whether my participants actually perceived it that way while playing the CDG, I asked them to assess their own behavior within the experiment in my post-experimental questionnaire.[75] On this basis, in appendix B.3., I analyze how consistent my interpretation of both players' behavior is with their own assessment of it. Based on their answers, I find that the larger the advisors' extent of lying, the more dishonest they perceived their own behavior in the game (*see Finding B-2 in appendix B.3.*). This demonstrates that the self-perception of the advisors indeed depends on the size of the lie. In addition, my results reveal that the participants considered truth-telling and cooperative behavior in fact as honest, while considering selfish lying as dishonest. This is consistent with my taxonomy of lies and truth-telling. Interestingly, more dishonest behavior was also associated with a higher preference for risk (*see Finding B-1*). Turning to the investors, I find that the higher their extent of mistrust, the more mistrusting the investors rated their own behavior in the game (*see Finding B-4*). From this it follows that the self-perception of the investors depends on the size of their mistrust. Moreover, my results show that the investors perceived risk-seeking mistrust in the game in fact as risk-seeking and risk-reducing mistrust as risk-averse, which speaks in favor of this terminology (*see Finding B-3*). Finally, my participants considered the act of following their received advice indeed as trusting. These findings are consistent with my taxonomy of mistrust and trust. It can be concluded that the size of the lie and the size of mistrust do not only matter from a strategic perspective but also have an impact on how people perceive their own behavior.

In the next section, I will discuss these findings against the backdrop of the existing literature, including possible drivers and inhibitors of lying and mistrusting behavior in strategic deception.

8. Discussion

The purpose of this paper was to provide a novel experimental design that allows measuring lying and mistrusting behavior on continuous and easily comparable scales. For this purpose, I designed a new sender-receiver game: the *Continuous Deception Game* (CDG). This experiment allows observation of lying, mistrust and respective first-order beliefs on continuous scales. The additional information that is gained by observing the extents (rather than the frequencies) of lying and mistrust is essential to understanding how strategic deception works, not only in this experiment. For instance, the proportion of investors who suspected their advisors to have lied was identical to the proportion of advisors who actually lied.[76] Solely based on this information, one might conclude that the investors' beliefs about their advisors' dishonesty were highly accurate. However, considering the size of the lie reveals that the investors strongly underestimated their advisors' extent of lying. In fact, the advisors overstated the true value of the given optimum by about 148% on average, while the investors suspected them to do so by only 56%. This misjudgment resulted in the investors relying too much on their received advice and therefore overinvesting to a high extent (while wrongly expecting to make near-optimal investments).

[75] In this questionnaire, I asked, on the one hand, the advisors to rate their preference for honesty and their preference for risk on a 7-point-scale. On the other hand, I asked the investors to rate their preference for trust and their preference for risk on a 7-point-scale. More details on this can be found in appendix B.3.

[76] Both proportions were approximately 78%.

This example highlights the importance of including the size of the lie and the size of mistrust into the picture when one aims to understand how dishonest or mistrusting people actually behave.

For this reason, studies that observe lying and mistrust as discrete, or even dichotomous, variables might yield completely different results if players were offered to choose to what extent they want to engage in dishonest or mistrusting behavior. With that in mind, I will discuss the meaning of my findings in the light of the existing literature, firstly, for the advisors (*8.1.*) and, secondly, for the investors (*8.2.*).

8.1. Lying and expected mistrust

There are many reasons why people lie or tell the truth. One of them is their expected *monetary gain*. On that matter, Gneezy (2005) finds in his sender-receiver game that people are sensitive to their monetary gain when deciding to lie. In particular, his results show that an increase of the profit the senders can expect from lying leads to a raise in the proportion of senders who lie – even if the losses that their lies are expected to cause to the receivers are increased by the same extent. Whereas Gneezy (2005) varies possible gains and losses from lying between treatments, in my experiment the senders (i.e., the advisors) can decide on the size of their lies (i.e., their extent of lying) themselves. In addition, gains and losses from lying in my experiment increase simultaneously with the extent to which the receivers (i.e., the investors) follow misleading advice.[77] Transferring Gneezy's (2005) findings to my experiment would suggest that the likelihood that a strategy which involves lying is pursued by the advisors increases with the expected profits associated with that strategy. This would also be in line with Fischbacher and Föllmi-Heusi (2013) and Gneezy et al. (2018) who find that most of their participants chose payoff-maximizing lies over partial ones when reporting the result of a die roll. My results however provide another perspective: If people can freely decide on the extent to which they want to lie, they rarely lie to the fullest possible extent (i.e., the players in my experiment do so in less than nine percent of all cases). Most advisors told selfish lies but they seldom fully exhausted their possibilities to lie. These results are consistent with a concept of lie aversion based on moral costs of lying that increase monotonously with the size of the lie (Gneezy et al., 2018; Lundquist et al., 2009). They are also in line with the aim to maintain a favorable self-concept (Mazar et al., 2008). Beyond that, even though the participants were anonymous to the experimenters, the advisors could have also been concerned about the experimenters' ex post judgment of their dishonesty (Utikal & Fischbacher, 2013).

While my paper does not aim to decide in favor of or against one of these theories (in fact, with the right conceptualization, all of them are in line with my findings), it offers another explanation: Strategic deception certainly involves expectations towards the own ability to manipulate others when lying. In a

[77] It should be reminded that Gneezy (2005) and I use the same ratio between the senders' profits from lying and the associated costs to the receivers. When Gneezy (2005) increases the profits to the senders, he raises the losses to the receivers by the same amount. Thus, the profit-loss ratio between his respective treatments is equal to 1. In my experimental design, the ratio to which gains and losses from lying are expected to increase depends on the ratio between both players' payoff parameters (m_A and m_I). Since, in this paper, I use payoff parameters with equal values ($m_A = m_I$), the resulting profit-loss ratio is identical to that used by Gneezy (2005).

context with minimal social interaction, these expectations are condensed in beliefs about another person's trusting behavior. More precisely, they are result-oriented beliefs a potential liar holds about how effective he or she can manipulate another person into following a desired course of action. On this matter, I find that the advisors in my experiment believed that giving higher advice above a certain level would not lead to higher overinvestments, since they expected their investors to disproportionally mistrust particularly high advice. As my results speak in favor of highly strategic and belief-based decision making, this indicates that some advisors thought that lying to a higher extent than they already did would not deceive their investors any further and, thus, not yield higher payoffs. In other words, my findings suggest that people who refrain from lying to the fullest possible extent might still lie to the highest extent they expect to be convincing. Such considerations about the own *potential to manipulate* in strategic deception cannot be addressed in most other economic experiments, since lies in other experiments often do not have to be convincing in order to yield favorable results for the liar (e.g., Fischbacher & Föllmi-Heusi, 2013; Mazar et al., 2008; Utikal & Fischbacher, 2013).[78]

One other experimental design in which such considerations certainly matter (even though they are not in the direct focus of the respective paper) is Lundquist et al.'s (2009) sender-receiver game. In their experiment, the senders were financially incentivized to lie about their individual test score in case it was below a certain threshold. To gain from lying, they needed to convince their receivers to sign a fixed-payment contract, which was only beneficial to the receivers if the senders had a test score equal or superior to the given threshold. Due to this binary payoff structure, the sizes of the senders' lies did not matter beyond the fact whether they convinced their receivers to sign the contract or not. Therefore, in contrast to my experiment, the senders were not incentivized to convince the receivers that they had the highest possible score. Under these conditions, Lundquist et al. (2009) find that none of the senders lied to the fullest possible extent. However, they observe a great fraction of lies noticeably above the given threshold. These results cannot be explained by costs of lying that increase with the size of the lie alone. Without any conceptualization of the fact that, in order to be successful, deception needs to be convincing, there would be no need to lie to an unnecessarily high extent. Following this line of argumentation, my findings on belief-based considerations about one's potential to manipulate another person provide a reasonable explanation for their results – namely that the senders in Lundquist et al.'s (2009) experiment might not have believed that the receivers would trust them, if they claimed to have test scores that were too close to the given threshold or the highest possible score. It can be concluded that, when liars need to convince others of their honesty, the extent of lying that is expected to maximize their payoffs does not necessarily correspond to the fullest possible extent of lying. While this

[78] I argue that this also applies to all sender-receiver games that feature binary-like choices (e.g., Dreber & Johannesson, 2008; Erat & Gneezy, 2012; Gneezy, 2005; Gneezy et al., 2020; Jacquemet, Luchini, Rosaz, et al., 2019; Jacquemet et al., 2018; López-Pérez & Spiegelman, 2013; Peeters et al., 2013, 2015; Sánchez-Pagés & Vorsatz, 2007; Sutter, 2009; Vranceanu & Dubart, 2019). While such experiments may allow for sophisticated deception through truth-telling (as shown by Sutter, 2009), considerations about the own potential to manipulate the other player are still strictly limited by binary-like strategy sets. By contrast, my results suggest that such considerations are based on the size of the lie and the size of expected mistrust.

demonstrates that sophisticated liars use the size of the lie as an instrument to manipulate others in strategic deception, it also shows the importance of considering this aspect when comparing different-sized lies in particular and when analyzing the intent behind lying or any other deceptive behavior in general.[79]

Of course, people who lie are not solely considering their potential monetary gain from lying. In fact, there are many other different drivers and inhibitors of deceptive behavior, which all account for different types of lying and truth-telling. To begin with, Erat and Gneezy (2012) argue that absent costs of lying, one would expect people to always tell profitable white lies (i.e., lies that constitute a Pareto improvement). However, in their Deception Game, they provide evidence that a significant fraction of senders tells the truth when offered the binary choice between truth-telling and telling such a lie. Therefore, they suggest that at least part of the reason why people tell the truth may be connected with some form of *intrinsic costs of lying*. In support of this, I observed a similar type of unprofitable truth-telling. Beyond that, I find that not all players who engaged in profitable white lies did so to the fullest possible extent. Thus, these players lied but did not exhaust the full potential of expected Pareto improvement. That is interesting, since this cannot be explained by lie aversion that does not consider the extent of lying (as suggested by Hurkens & Kartik, 2009). This indicates that the respective players were dealing with moral costs of lying that increase with the size of the lie (as suggested by Gneezy et al., 2018; Lundquist et al., 2009).

Moreover, I can add to this matter that the proportion of unprofitable truth-telling diminishes to less than five percent when the players can freely decide about their extent of lying. On the one hand, the mere existence of such behavior speaks in favor of some *pure lie aversion* (as suggested by Erat & Gneezy, 2012), which is in contrast to Vanberg (2017) who finds no evidence for the existence of such a motivation in his experiment. On the other hand, the small fraction of unprofitable truth-tellers implies that for most players lying at least to a small extent was acceptable when they expected that doing so

[79] Gneezy et al. (2018) provide yet another explanation for why expectations about the own credibility in front of others are important. They show that potential liars care about their *social identity*. This concept captures concerns the subject has about how he or she is perceived by others. Since these concerns refer to beliefs one has about another person's beliefs about oneself, these concerns are based on second-order beliefs. By contrast, strategic considerations about one's potential to manipulate others focus on beliefs one holds about how effective one can trick another person into choosing a desired course of action. Hence, such considerations are based on simpler first-order beliefs. In my experiment, I asked players solely for their first-order beliefs, since there is evidence that first-order beliefs already sufficiently capture the relation between beliefs and behavior in sender-receiver games, whereas second-order beliefs do not provide much more insight (López-Pérez & Spiegelman, 2013). While this allows keeping my experiment simpler, it makes it hard to ex post distinguish social identity concerns (Gneezy et al., 2018) from strategic considerations about one's potential to manipulate others. To make this distinction, one could conduct a version of the CDG in which players are asked for their second-order beliefs in addition to their first-order beliefs.
Since my experiment involves a minimum of *social interaction* (which is similar to many other experimental designs, such as those of Erat & Gneezy, 2012; Fischbacher & Föllmi-Heusi, 2013; Gneezy, 2005; Mazar et al., 2008; Pruckner & Sausgruber, 2013; Sutter, 2009), it could also be interesting to implement different ways of communication between the advisors and investors. In combination with elicited first- and second-order beliefs, this would allow analyzing the impact of social interaction on social identity concerns and strategic considerations about one's potential to manipulate others.

would yield a Pareto improvement. This in turn is in support of Vanberg (2017), as this means that intrinsic lie aversion alone seems to not have been a strong driver for absolute truth-telling in my experiment.

But then, what made players tell the truth? My results reveal that truth-telling is four times more likely to happen when players expect to be trusted by their co-players. In addition, the extent of lying decreases with a decreasing extent of expected mistrust. These findings suggest that both truth-telling and lie aversion are closely connected to the *expectation of being trusted*, which indicates that people want their honesty to be rewarded with trust.

Truth-telling is generally seen in a more positive light than telling a lie. While this assessment is certainly true in most cases, not all lies are of bad intentions. For instance, in Erat and Gneezy's (2012) sender-receiver game a significant fraction of senders chose to engage in altruistic white lies (i.e., lies that are expected to help others at the expense of the liar) when offered the binary choice between telling such a lie or the truth. However, my results reveal that *altruism* is not a strong driver when the players can decide on the sizes of their lies themselves, since almost none of the subjects in my experiment (i.e., less than one percent) told this type of lie. Instead most advisors engaged in behavior that was expected to yield higher payoffs for themselves, such as selfish lying. This shows that altruism can be heavily undermined by other motivational factors, such as monetary gain.

According to Erat and Gneezy (2012), selfish lies (i.e., lies that help the liar at the expense of another) are expected to evoke *guilt*, whereas profitable white lies (i.e., lies that constitute a Pareto improvement) are not. In line with this, Erat and Gneezy (2012) find that the fraction of senders who lie is significantly higher for profitable white lies than for selfish lies. When interpreting their results, one must remember that the senders in Erat and Gneezy's (2012) experiment could never actively choose between these two types of lying. However, what happens if people can financially benefit from turning a profitable white lie into a selfish lie by further increasing their extent of lying? My results show that, when given this choice, most advisors chose selfish over profitable white lies. In fact, I observed nearly three times as many selfish lies as profitable white lies, which implies that, in most cases, neither additional costs of lying nor potential feelings of guilt were able to prevent players from telling selfish instead of profitable white lies. That difference between my results and those of Erat and Gneezy (2012) illustrates that the intention to tell a lie that is not only beneficial to the liar but also helps another person can be crowded out by adding the possibility of additional gain through telling a selfish lie. This suggests that people who tell lies that are mutually beneficial to themselves and to another person might not really care about the other person but use the fact that they are helping that person to rationalize the lie (Beck et al., 2020).

Overall, the advisors in my experiment seem to have been largely driven by strategic considerations about how to increase their monetary gain, while altruism and guilt appear to have barely held them back from opportunistic behavior. However, my findings suggest that there is some form of intrinsic lie aversion which appears to increase with the size of the lie. Finally, it seems that the expectation of being trusted can foster completely honest behavior.

8.2. Mistrust and suspected lying

The findings that I have discussed so far have shown the importance of trust and mistrust to the advisors. Many other studies on lying and truth-telling, however, examine lying behavior detached from trust (e.g., Fischbacher & Föllmi-Heusi, 2013; Mazar et al., 2008; Pruckner & Sausgruber, 2013) or at least focus more on the former than on the latter (e.g., Erat & Gneezy, 2012; Gneezy, 2005; López-Pérez & Spiegelman, 2013; Lundquist et al., 2009; Vranceanu & Dubart, 2019). While this focus on lying behavior serves the purpose of these studies in a good way, it also ignores the importance of trust for economic decision making. To address this issue, my experiments allows drawing conclusions on why people trust or mistrust potential liars.[80]

On this basis, my results reveal four main *drivers of (mis)trusting behavior*: monetary gain, risk aversion, altruism, and some form of endogenous preference for trust. In addition, they suggest that trust is the result of strategic and belief-based decision making. This is consistent with studies that show that trust in both a sender-receiver game with binary choices (Peeters et al., 2015) and the Trust Game (Sapienza et al., 2013) is based on first-order beliefs. I can add to this matter that the extent of one's mistrusting behavior is strongly connected to the first-order beliefs one holds about the extent of another person's lying behavior. In particular, the extent of mistrusting behavior in my experiment increases with the extent of suspected lying. As a result, players engaged nearly six times more often in completely trusting behavior when they expected their co-players to have told the truth. Thus, the expectation of being told the truth seems to be a strong driver for cooperative trust, which is consistent with the motive of expected payoff maximization in my experiment. Even though a major fraction of investors suspected their advisors to have lied to them, most of them still used their received advice as an important reference point for their investments, which still indicates a general inclination to trust. However, most of the investors who suspected their received pieces of advice to be lies engaged in mistrusting behavior to improve the quality of their investments by (at least partially) compensating the extent to which they expected their advisors to have lied. In this way, they raised their expected payoffs. For this reason, I refer to such behavior as profitable mistrust, or rather profitable white mistrust if it also increased the advisors' payoffs. Similar to cooperative trust, profitable (white) mistrust could be motivated by *monetary gain*.

[80] It should be reminded that, in my experiment, *trust* refers to honesty-related trusting behavior (i.e., one's reliance on another person's honesty), whereas *trusting beliefs* correspond to expectations about the honesty of another person. Note that my understanding of trust in the CDG differs from that in the original version of the Trust Game in which trust refers to the act of relying on another person's social cooperation (instead of on another person's honesty).

Apart from this, a significant fraction of investors engaged in excessive mistrust, which cannot be explained by expected payoff maximization. In these cases, the investors invested less than they expected to be optimal. While they could expect this to reduce both players' payoffs, they could also expect it to decrease the risk associated with their own payoffs, since lower investments hold a lower risk for the investors in my experiment by design.[81] This suggests that, to some extent, the investors valued *risk aversion* over monetary gain and, therefore, engaged in more mistrusting behavior. In support of this, the investors' ex post self-assessment of their preference for risk within the experiment is largely consistent with the risk that is inherent to their displayed mistrust in the game. It can be concluded that risk aversion can be another driver of mistrusting behavior. This is in line with Sapienza et al. (2013) who find that trust in the Trust Game is correlated with a preference for risk tolerance. However, my findings add that this applies not only to trust that refers to expectations of social cooperation (as in the Trust Game) but also to honesty-related trust (as in the CDG).

Furthermore, I observed benevolent trust and benevolent mistrust. In these cases, the investors expected their (mis)trusting behavior to result in overinvestments, which would reduce their payoffs while increasing the payoffs of the advisors. This type of behavior can neither be explained by expected payoff maximization nor by risk aversion (on the contrary, benevolent mistrust increases the risk associated with the investors' payoff by design). One explanation for such behavior could be that the respective players valued risk-taking over monetary gain. However, the rates at which players engaged in benevolent (mis)trust are not correlated with their self-assessed risk preference. This speaks in favor of another explanation, namely that the respective players had some preferences over distributions of payoffs and, by intentionally making overinvestments, aimed to increase their co-players' payoffs. Following this line of reasoning, it is plausible to assume that *altruism* was another driver of (mis)trusting behavior in my experiment. These findings complement those of Sapienza et al. (2013) who observe a similar pattern for trusting behavior in the Trust Game. In addition, my findings are consistent with Cox (2004) and Innocenti and Pazienza (2006) who use a clever triadic design to show that altruism is one decisive factor for trusting behavior in the Trust Game. Beyond the scope of the cited studies, my findings provide evidence that not only trusting but also mistrusting behavior can be driven by altruism.

Finally, I observed that some investors engaged in unprofitable trusting behavior (i.e., even though they believed that their advisors had lied by understating the optimal investment, they still exactly followed their received advice). As a result, they expected to make underinvestments, which would reduce both players' payoffs when compared to more mistrusting behavior. Since such unprofitable trust is expected to yield a Pareto deterioration, monetary gain and altruism can be ruled out as the sole driving factors behind this type of trusting behavior. One could argue, though, that such investments were the result of

[81] It bears repeating that, in the CDG, the size of a completely risk-reducing investment is zero. From that point upwards, the inherent risk of the investment increases with the size of the investment by design.

a trade-off between monetary gain (or altruism) and risk aversion. If this was the case, however, it would be improbable that the investments of the respective players would correspond exactly to their received pieces of advice (which they believed to be lies anyway). Thus, a combination of monetary gain, altruism, and risk aversion cannot fully explain the observed fraction of unprofitable trusting behavior. Moreover, it can be excluded that investors who engaged in unprofitable trusting behavior wanted to reward honesty with trust, since they expected their advisors to have lied. On this basis, I argue that the respective players assigned a positive value to trusting behavior per se. This suggests that people may have some *endogenous preference for trust*, which would be consistent with studies that link trust to positive (Barefoot et al., 1998; Kuroki, 2011) and mistrust to negative sensations (Gurtman, 1992).

In the next and last section, I will summarize my most important findings.

9. Conclusions

By introducing continuous variables to the sender-receiver game, the *Continuous Deception Game* (CDG) enables the measurement of the extents of lying and mistrusting behavior as well as both players' first-order beliefs on continuous scales. Due to the resulting continuous message space, this experiment can address the issue of sophisticated deception through truth-telling (Sutter, 2009). Beyond that, it allows researchers to make other types of sophisticated deception (such as the extent to which a lie is expected to manipulate another person) visible. Therefore, it enables distinctions to be made among a broad range of strategies for both players. By way of this method, the CDG sheds new light on several aspects of lying and mistrust in strategic deception.

In the first place, with regard to *lying* and *truth-telling*, my results are in support of lying costs that increase with the size of the lie. However, I find only weak evidence for pure lie aversion. In addition, it seems that, when people can decide on their extent of lying themselves, altruism and guilt aversion can be largely undermined by the possibility of additional monetary gain that results from lying to a larger extent. Comparing these findings to those of Erat and Gneezy (2012) suggests that people use altruistic motives to rationalize lying for selfish reasons. Moreover, my findings indicate that people make strategic considerations about their own potential to manipulate others based on the size of the lie. In particular, sophisticated liars anticipate that lies that are of an unrealistically high extent (or, in other words, are too close to the fullest possible extent of lying) will be disproportionally mistrusted and therefore fail to further manipulate their recipients. Finally, I find evidence that people behave more honestly when they expect their honesty to be rewarded with trust.

In the second place, I can identify four main drivers of *trusting* and *mistrusting behavior*. To begin with, I find that people might have some endogenous preference for trust. In addition, I provide evidence that mistrust can be motivated by expectations of additional monetary gain, as well as by excessive risk aversion, for which the decision makers are even willing to accept a reduction in their expected payoffs.

Lastly, my results indicate that, when mistrusting behavior can actually help the mistrusted person, mistrust can also be driven by altruism.

In conclusion, there is a wide spectrum of internal and external factors that can drive (dis)honest and (mis)trusting behavior – and a broad range of them can be analyzed in the CDG. This demonstrates the variety of application possibilities of this experiment and shows its potential as a straightforward method to analyze the relationship between honesty, trust, and respective beliefs in strategic deception.

CHAPTER 4

Quick, intuitive truth-telling and deliberate, strategic lying under oath

Tobias Beck[♣]
Institute of Economics, University of Kassel, Nora Platiel-Straße 4, 34127 Kassel, Germany

Abstract. Ethical oaths are frequently used as a management device to foster honest behavior. Even though most executive decisions involve strategic interaction, the mechanisms behind oath-taking in strategic interaction are not yet fully understood. Against this backdrop, I analyze how oath-taking works when lying requires strategic thinking. Therefore, I design a laboratory experiment to test how an honesty oath affects deliberation time, strategic decision making, and the temporal consistency of (dis)honest behavior. On the upside, my results show that an honesty oath leads to more truth-telling by making the decision to tell the truth less deliberate and more intuitive. On the downside, I observe that people who lie under oath barely reduce the sizes of their lies. Here, I find that the act of lying becomes more deliberate and strategic due to the oath. Finally, I provide evidence that oath-taking makes both the oath-takers' lying and truth-telling behavior more consistent over time. This implies that an honesty oath encourages the oath-takers to make one basic decision whether to commit to unconditional truth-telling or to deliberately accept being dishonest for personal financial gain.

Keywords: Honesty oath, Strategic deception, Truth-telling, Size of the lie, Laboratory experiment.

JEL: C91, D03, D63.

Availability of data: The datasets generated and analyzed in this paper are available from the author of this dissertation on reasonable request.

Informed consent: Informed consent was obtained from all individual participants included in the study.

[♣] Email: tobi.beck@gmx.net; ORCID: 0000-0002-4162-3648

1. Introduction

In the face of financial crises, public officials and regulatory authorities proposed various measures to enhance commitment to ethical standards in the business world. One proposal is that business leaders and public officials should be obliged to take ethical oaths when coming into office. In recent years, ethical oaths were implemented for both entire vocational fields (e.g., the MBA Oath) and entire professions (e.g., the Dutch Banker's Oath). In the course of this development, oath-taking for senior executives as a means to raise commitment to moral standards and to honest behavior has become more prominent in the last decade (de Bruin, 2016; Jacquemet, Luchini, Rosaz, et al., 2019). As a matter of fact, senior executives and employees in some businesses are already obliged to take honesty oaths (de Bruin, 2016). Against this backdrop, it is of utmost importance to understand whether and how such oaths work and which effects they have on the oath-takers. While there is broad empirical evidence that honesty oaths work at least sometimes in terms of inducing behavioral changes (see, for example, Beck et al., 2020; Carlsson et al., 2013; Jacquemet et al., 2013; Kemper et al., 2016), much less is known about the mechanisms behind them.

In this paper, I aim to get a better understanding of *how* honesty oaths work. I implement a between-subjects design with repeated measurements to examine the causal effects of an honesty oath on moral deliberation, strategic decision making, and the temporal consistency of (dis)honest behavior. Therefore, I use Beck's (2020) Continuous Deception Game (CDG), which introduces continuous variables to the sender-receiver game. This allows for observation of more differentiated lying behavior (i.e., the size of the lie) and first-order beliefs (i.e., the extent of expected mistrust). This is important because lying behavior in previous studies about behavioral effects of honesty oaths is often observed as the much simpler binary choice between truth-telling and lying. For this reason, there has been little focus on the impact of oath-taking on the size of the lie.[82] With that in mind, in this study, I analyze how an honesty oath affects the extent to which liars lie. Moreover, I observe the decision time, which is used in standard literature to assess the extent of cognitive reasoning and moral deliberation in the decision-making process (e.g., Jacquemet, Luchini, Rosaz, et al., 2019; Kahneman, 2011; Rubinstein, 2007; Schunk & Betsch, 2006). In addition, I analyze how the oath influences the relationship between the displayed lying behavior and beliefs about others. This allows drawing conclusions on the impact of the oath on belief-driven strategic decision making.

I find that the honesty oath increases the overall truth ratio but has no significant effect on the size of the lie of people who lie even under oath. In fact, liars under oath tend to deliberate more on their decisions and behave more strategically, whereas the additional truth-tellers under oath tend to deliberate less and make more intuitive decisions. Overall, having subjects sign an honesty oath makes their lying and truth-telling behavior more consistent over time, which indicates that the oath aligns their behavior

[82] One of the few studies, which address this issue, are Beck et al. (2020) and Jacquemet et al. (2020).

to some internal reference point of what is right or wrong in the given situation. By this means, I provide evidence of how the honesty oath performs when taking it is not a free choice but a precondition for participating in the experiment. This is important because some ethical oaths in the business world such as, for instance, the Dutch Banker's Oath are also not made voluntarily but are a legal requirement.

I begin by giving a short review of the relevant literature in section 2. It follows an explanation of my experimental design and treatments in section 3. On this basis, I derive hypotheses about the effects of the honesty oath in section 4. Subsequently, I provide details on my experimental procedure in section 5. Afterwards, I report the results of my experiment in section 6, which follows the structure of the hypotheses section. Finally, I discuss my most important findings in the light of the existing literature in section 7.

2. Literature review

2.1. Ethical oaths and their application

Oaths have existed for thousands of years. According to Boatright (2013), they even precede religion, since they are based on primitive beliefs about the power of curses to harm others. In modern days, they are still used in various areas. For instance, public officials are sworn when they are given the status of civil servants. But also, in the private sector, many companies implement ethical clauses in their codes of conduct in order to foster honest behavior and to grow moral commitment among their employees. Some oaths are also associated with entire professions, including medicine and law. One much discussed example is the Dutch Banker's Oath, which originated from the financial crisis that began in 2007. This oath aims at promoting ethical banking practices among bankers by focusing their attention on shareholder interests. All Dutch bank employees are legally required to take this oath. By this means, the oath attempts "to restore trust in an industry whose reputation was tarnished by the recent financial crisis" (Boatright, 2013, p. 161). Another well-known oath, which aims at fostering ethical behavior, is the MBA Oath. This oath originated from the Harvard Business School in 2009 and was inspired by the Columbia Business School's honor code. It is voluntary and primarily directed to MBA students and graduates. By taking that oath, one pledges to behave ethically and responsibly in exercising one's professional duties. Anderson and Escher (2010) state that the MBA Oath intends to raise commitment to approaching ethically important decisions thoughtfully and consciously. The same principle applies to ethical oaths in general: They are used as devices to guide the oath-takers' behavior in a desired direction.

2.2. Overall effects of honesty oaths

As ethical oaths have been implemented more frequently in the private sector in recent years, the question of whether and how they affect the oath-takers has become an eminent topic of research. Blok (2013) argues that an oath changes the oath-taker's identity by manifesting their intention “not only to *do* something, but also to *be* the one who is committed to some future course of action” (Blok, 2013, p. 193, italics in original). This is consistent with Abeler et al. (2019) who find in a metastudy, which is based on 90 empirical studies from economics, psychology, and sociology, that being honest is one of the main motivations for truth-telling. The implied superiority of intrinsic over extrinsic motivation to behave honestly is supported by Pruckner and Sausgruber (2013) who show that appealing to honesty can mitigate dishonest behavior more effectively than a reminder of legal norms. In another study, Mazar et al. (2008) find that one’s tolerance for dishonest behavior can be reduced by directing one’s attention to moral standards. They suggest that people who could gain financially from cheating at the expense of an honest self-concept try to find a balance between these two motivational forces. Based on that, Mazar et al. (2008) argue that a moral reminder can decrease people’s self-concept maintenance thresholds and, therefore, results in more honest behavior. In line with this, Koessler et al. (2019) show that the compliance to pay taxes on time can be increased by a mere promise to do so. Both in field and laboratory experiments, they find that a non-financial reward is more likely to generate positive commitment than a financial one.

There is a wide range of evidence that ethical and, in particular, honesty oaths are able to foster honest behavior: Carlsson et al. (2013), Jacquemet et al. (2013), as well as Kemper et al. (2016) find that an oath can increase the honesty of survey respondents who are asked for their hypothetical willingness-to-pay. As demonstrated by Jacquemet et al. (2018), an honesty oath can also lead to more truth-telling in a modified version of Rosenthal’s (1981) sender-receiver-game. This is supported by Jacquemet, Luchini, Rosaz, et al. (2019) who find similar effects in Erat and Gneezy’s (2012) sender-receiver game. In addition, Beck et al. (2020) provide evidence that an honesty oath can even mitigate lies in disguise as introduced in the dice experiment by Fischbacher and Föllmi-Heusi (2013). As Beck et al.’s (2020) laboratory experiment neither features financial nor social punishment, they argue that the honesty oath mitigates lying by increasing the oath-takers’ moral awareness. This is supported by de Bruin (2016) who suggests that oaths may facilitate moral deliberation.

While there is little doubt about the general ability of honesty oaths to enhance truth-telling, less is known about their impact on the oath-takers’ decision-making process. I will address this issue in the next subsection.

2.3. Moral deliberation and strategic decision making under oath

Decision time is often used to assess the extent of cognitive reasoning in the decision-making process (e.g., Jacquemet, Luchini, Rosaz, et al., 2019; Kahneman, 2011; Rubinstein, 2007; Schunk & Betsch, 2006). In a basic model of the cognitive system that is used in decision making, Kahneman (2003) distinguishes “between two generic modes of cognitive function, corresponding roughly to intuition and reasoning” (Kahneman, 2003, p. 1450), where reasoning is more time-intensive than intuition. This is supported by Rubinstein (2007) who finds that the decision time is associated with the extent of cognitive reasoning in decision making, indicating that longer decision times can lead to more strategic decisions. In line with this, Schunk and Betsch (2006) show that the more intuitive a subject is, the faster the subject completes different lottery tasks. These findings suggest that strategic decision making takes more time than intuitive decision making. In addition, there is empirical evidence that the decision time increases with the extent of conflict that subjects feel about their decisions in the public goods game (Evans et al., 2015). Thus, the decision time seems also to depend on the inner conflict of the decision maker. Moreover, Capraro et al. (2019) find that time pressure, i.e., a forced reduction of decision time, increases honesty in a variation of Gneezy's (2005) sender-receiver game, which suggests that truth-telling is more intuitive than lying.

With that in mind, how do honesty oaths affect the oath-takers' decision-making process? Jacquemet, Luchini, Rosaz, et al. (2019) show that taking an honesty oath prior to participation in Erat and Gneezy's (2012) sender-receiver game can both increase the decision time of liars in a neutral environment (when a lie is not called a lie) and reduce the decision time of truth-tellers in a loaded environment (when a lie is called a lie). They explain these findings by assuming that the honesty oath strengthens the moral dilemma in the game and, therefore, eases truth-telling while making lying more difficult. In another study, Jacquemet, Luchini, Shogren, et al. (2019) introduce an honesty oath and two other oaths that aim at fostering cognitive effort to a discrete choice experiment, in which players can identify payoff-maximizing choices by calculating the difference between the value and the costs of given tokens. Jacquemet, Luchini, Shogren, et al. (2019) find that all three oaths increase the decision time but that only the honesty oath improves the quality of decision making by increasing the proportion of payoff-maximizing choices. Based on this, they argue that the honesty oath does not foster cognitive reasoning but the effort to tell the truth, which requires more time from the subjects to identify the payoff-maximizing option. Thus, Jacquemet, Luchini, Shogren, et al. (2019) suggest that the honesty oath reduces the tendency to be strategic.

Another interesting point is made by Hergueux et al. (2019) whose results suggest that individuals behave more consistently with their internalized social norms in the public goods game when they have signed an honesty oath. In particular, they find that an honesty oath "works mainly by inducing subjects to put more weight on internal norms rather than strategic thinking" (Hergueux et al., 2019, p. 15). This is supported by the fact that they observe a decrease in decision time under oath. Moreover, they observe no change in the proportion of subjects who declared thinking about the contributions of others while making their own decision. On this basis, they argue that the behavioral change under oath is not due to belief-driven strategic reasoning. They conclude that the honesty oath leads to less strategic and more intuitive decision making.

As demonstrated by these last two studies, one challenge when interpreting the effects of honesty oaths on the decision time is to distinguish moral deliberation from strategic reasoning. The analysis of whether and how honesty oaths impact strategic considerations when lying requires assessment of the extent to which lying and truth-telling result from strategic decision making. Against this backdrop, it should be noted that, in general, strategic decision making is based on beliefs about other people. The same seems to apply to strategic deception, as there is evidence that lying behavior in a strategic environment is also based on beliefs about others. For instance, López-Pérez and Spiegelman (2013) show that the first- and second-order beliefs of the senders in the sender-receiver game are correlated with their decision to lie. This is in line with Lundquist et al. (2009) who provide evidence that the senders' propensity to lie in the sender-receiver game is strongly associated with their beliefs about deceptive lies. In another sender-receiver game, Peeters et al. (2015) find that the senders base their decisions whether to tell the truth or not on their first-order beliefs about the receivers. More precisely, Peeters et al. (2015) observe that the senders lie less often if they expect to be mistrusted by the receivers, which suggests that they have a tendency for strategic deception (Sutter, 2009).[83] All these studies indicate that strategic reasoning in deciding whether to lie or not is reflected in a strong relationship between the displayed lying behavior and beliefs about others. Little is known, however, about how an honesty oath impacts this belief-behavior relationship. Thus, analyzing the potential effects of oath-taking on this belief-behavior relationship would allow drawing conclusions on whether changes in decision time under oath are associated with more or less strategic decision making. This is where this paper aims to contribute.

3. Experimental design

3.1. Baseline treatment (BT)

My baseline treatment is identical to Beck's (2020) Continuous Deception Game (CDG). This experimental design introduces continuous variables to Erat and Gneezy's (2012) sender-receiver game and, therefore, allows for observation of the individual extent of lying and of expected mistrust. The sender of this game (i.e., an advisor) has private information on the true value of an optimal investment, which maximizes the payoff of the receiver (i.e., an investor). This optimal investment is a random number between 0 and 100 and is determined hidden from the players before the game starts. The game consists of two stages: In the first stage, the sender is informed about the value of the optimal investment and instructed to report it to the receiver by choosing an advice number between 0 and 100. In the second stage, this piece of advice is passed to the receiver who then is instructed to make an investment by choosing a number between 0 and 100. This investment determines both players' payoffs. While the payoff of the sender increases with the size of the investment, the receiver's payoff decreases by any upward or downward deviation of the investment from the true optimal investment. Both the exact definition and an example of both players' payoff functions can be found in appendix C.1. With this

[83] In binary choice-based sender-receiver games (e.g., Peeters et al., 2015), senders who expect their receivers to mistrust them (by not following their message) may send the true message for precisely this reason (Sutter, 2009). Sutter (2009) refers to such belief-driven strategic behavior as sophisticated deception. This illustrates the importance of beliefs about others in strategic deception.

payoff structure, both players have conflicting incentives: Whereas the sender is incentivized to lie by overstating the true optimal investment, the receiver is incentivized to invest optimally and therefore must rely on the alleged optimum reported by the sender (Beck, 2020).

Moreover, in the CDG, both players' first-order beliefs are elicited (Beck, 2020). On the one hand, the sender is instructed to estimate the size of the investment that they expect the receiver to make after receiving the advice number. On the other hand, the receiver is instructed to estimate the true value of the optimal investment.[84] Finally, to assess the extent of cognitive reasoning in decision making, I observe the decision time of the senders. In line with Jacquemet, Luchini, Rosaz, et al. (2019), I follow a standard definition in the literature by defining decision time as the time elapsed between the moment the concrete decision problem is presented to the player and the moment the decision is made.

Consistent with Beck (2020), I define the individual *extent of lying* as the percentage extent to which the sender's advice number deviates from the true optimal investment. In addition, I define the individual *extent of expected mistrust* as the negative percentage extent to which the sender expects the receiver's investment to deviate from the reported advice number (Beck, 2020).[85] Moreover, I define the *truth ratio* as the proportion of cases in which the senders report the true optimal investment to the receivers. Note that this paper focuses mainly on the behavior and beliefs of the senders. However, the receivers are needed in the game to encourage the senders to engage in strategic decision making. As a result, the receivers play a rather passive role in my experiment.

My experiment consists of ten rounds of the CDG. Before the start of the first round, every player is randomly assigned to one of the two roles (i.e., sender or receiver), which does not change throughout the entire experiment. In every round, each sender is randomly matched with a receiver. I use a perfect stranger design. Thus, players are re-matched between rounds such that none of them has the same co-player twice. This is known to all players. In addition, the players are informed about the outcome of all ten rounds of the CDG only after finishing the last round. This procedure rules out common learning effects and reduces interdependences between rounds (Beck, 2020).

The values of the optimal investment that I use in the ten rounds of the CDG are: 50, 70, 25, 35, 10, 50, 70, 25, 35, and 10 in this order. Note that, for simplification, I only use optimal investments that are divisible by five (Beck, 2020). In addition, each value is repeated with a time lag of five rounds. This ensures that, in the last five rounds, the senders are confronted with identical decision problems as in

[84] Both players' estimates are not monetarily incentivized because, according to Sutter (2009) who refers to Camerer and Hogarth (1999), "there is evidence that eliciting expectations with or without monetary rewards for accuracy does not yield significantly different results" (Sutter, 2009, p. 50). Moreover, I refrain from eliciting higher-order beliefs, since López-Pérez and Spiegelman (2013) find that first-order beliefs already capture the belief-behavior relationship in sender-receiver games, while "second-order beliefs do not appear to provide much more insight" (López-Pérez & Spiegelman, 2013, p. 243).

[85] Compared to Beck (2020), I invert the sign of the extent of expected mistrust. This is done because it can be expected that the receivers tend to make investments below their received advice numbers. Therefore, inverting the sign of this extent will make it easier to read statistical results later in this paper.

the first five rounds, which will be useful to examine the consistency of their lying behavior over time in more detail. Finally, to prevent players from basing their decisions on estimations of their average payoffs in previous rounds, the players are informed in advance that only one round will be selected randomly to determine their final payoff at the end of the experiment. Therefore, after one round is selected, their decisions in all other rounds do not count. This makes sure that the players have to decide in line with their preferences for lying for financial gain separately in each round.

3.2. Honesty oath treatment (HOT)

In my honesty oath treatment, each sender has to sign an honesty oath immediately after being assigned to their role at the very beginning of the experiment. Therefore, the oath is handed out to each sender individually on a separate sheet of paper by the experimenter.[86] The subject then has to read and sign the oath in front of the experimenter. Afterwards, they have to hand it back to the experimenter. In addition, the receivers in this treatment are informed about the exact wording of the oath, which in turn is known to the senders. The oath the subjects sign states:

> "*Hereby I do swear that my actions during this experiment will be due to the* ***principle of honesty***. *In particular, I swear not to lie in order to enrich myself.*"

This particular oath makes explicit that lying to enrich oneself is considered dishonest. As a result, the subjects could perceive this oath as more demanding than less concrete oaths such as, for instance, that used by Jacquemet, Luchini, Rosaz, et al. (2019) where subjects swore to "tell the truth and always provide honest answers" (p. 429). However, honesty oaths in practice are often even more demanding than the oath that I use, even if they do not explicitly address lying. For example, in U.S. courts, witnesses are required to swear that they "will tell the truth, the whole truth, and nothing but the truth" (Court Organization and Civil Code, 2013). Due to the social framework in which the court oath is embedded (Bachvarova, 2008), it is certainly perceived as more demanding than the oath in my treatment. In addition, the redundant formulation of the court oath makes absolutely clear that lying is not permitted, which reduces ambiguities that could be exploited to rationalize lying (Shalvi et al., 2015). There is a good reason not to copy the exact wording of the court oath in lying experiments[87] – yet, in order to avoid ambiguities in the formulation of the oath, I decided to use an oath that makes equally clear that lying is not permitted. The formulation that I use is inspired by the "Honor Code" of the Columbia Business School, which states: "[…] I adhere to the *principles of truth*, integrity and respect. I will *not lie*, cheat, steal, or tolerate those who do." (Columbia Business School, 2020, emphasis added).

[86] The honesty oath that I used can be found in appendix C.2.2.

[87] Note that certain oaths should be reserved for special occasions such as in court, since using them in non-special occasions could have a negative impact on the meaning that oaths currently hold in the public sector (Rutgers, 2010, 2013). To reduce such possible spillover effects, I refrained from using the exact wording of the court oath and instead used another oath that also makes clear that lying is not permitted.

While it has been argued that oaths are most effective when signed freely (Jacquemet et al., 2013), in practice not all oaths are made freely. If firms or banks make their senior executives take ethical oaths, the executives do not always have a choice but to take the oath (e.g., the Dutch Banker's Oath). The same applies to the aforementioned court oath or oaths of office, which public officials take when they are sworn into office (Rutgers, 2010). Here, voluntary oaths might be the better choice in terms of inducing behavioral change, as suggested by Jacquemet et al. (2013). However, obligatory oaths are typical both for the private and for the public sector. To mimic this, taking the oath in this treatment is not made a free choice but a precondition for participating in the experiment.[88] None of the subjects refused to take the oath when asked to sign it.

Apart from this honesty oath procedure, the HOT is implemented identically to the BT.

4. Hypotheses

4.1. Overall effects of the honesty oath

There is broad evidence that honesty oaths enhance truth-telling (see, for example, Beck et al., 2020; Hergueux et al., 2019; Jacquemet et al., 2018). I expect similar effects in my experiment, which leads to my first hypothesis:

Hypothesis H1. *The overall truth ratio is higher with an honesty oath (HOT) than without it (BT).*

However, little is known about the effect an honesty oath has on the size of the lie. On the one hand, Lundquist et al. (2009) provide evidence that the aversion to lie increases with the size of the lie and the strength of a promise not to lie. On the other hand, Jacquemet et al. (2020) show that only partial liars are affected by an honesty oath in an income/tax declaration experiment. They provide evidence that partial liars become fully honest after taking an honesty oath. In line with this, Beck et al. (2020) do not detect partial lies after introducing an honesty oath to Fischbacher and Föllmi-Heusi's (2013) dice experiment. These findings suggest that subjects who behave dishonestly after taking an honesty oath, might not reduce their extent of lying. While I expect that the average extent of lying in my experiment will decrease due to more complete truth-telling under oath, I predict that the honesty oath will not reduce the average extent to which liars lie. This leads to my next two hypotheses:

[88] In experiments that involve subjects taking oaths, there is always the risk of having experimenter demand effects (i.e., subjects could perceive an oath as a cue about the way in which they are expected to act during the experiment). Moreover, making an oath a precondition for participating in an experiment potentially increases that risk (for more details, see Jacquemet et al., 2013). As a result, there is a trade-off between reducing the risk of experimenter demand effects and using oath procedures that mimic common practices where oaths might be perceived as equally or even more demanding than in experiments. With that in mind, in this paper, I decided in favor of making the oath obligatory in order to recreate common practices. I owe this point to an anonymous referee.

Hypothesis H2. *The average percentage extent of lying is lower with an honesty oath (HOT) than without it (BT).*

Hypothesis H3. *The average percentage extent to which liars lie does not differ with an honesty oath (HOT) and without it (BT).*[89]

4.2. Quick truth-telling and deliberate lying under oath

In recent studies, it has been shown that an honesty oath can influence the oath-takers' decision time (Hergueux et al., 2019; Jacquemet, Luchini, Rosaz, et al., 2019; Jacquemet, Luchini, Shogren, et al., 2019). Based on these studies, I expect that the honesty oath will intensify the moral dilemma that liars face in my experiment. In addition, I expect that the oath will facilitate moral deliberation for truth-tellers, as suggested by de Bruin (2016). Under the assumption that the extent of cognitive reasoning and the perceived inner conflict in decision making are reflected in the decision time (see, for example, Evans et al., 2015; Kahneman, 2003, 2011; Rubinstein, 2007; Schunk & Betsch, 2006), I predict that the honesty oath in my experiment will increase the decision time of liars and reduce the decision time of truth-tellers. To test this, I propose the following pair of hypotheses:

Hypothesis H4a. *The average decision to lie takes longer with an honesty oath (HOT) than without it (BT).*

Hypothesis H4b. *The average decision to tell the truth takes shorter with an honesty oath (HOT) than without it (BT).*

4.3. Intuitive truth-telling and strategic lying under oath

There is evidence that strategic considerations when deciding about lying are reflected in the relationship between the displayed lying behavior and beliefs about others (see, for example, López-Pérez & Spiegelman, 2013; Lundquist et al., 2009; Peeters et al., 2015).[90] Based on the cited studies, I operationalize the extent of belief-driven strategic decision making in my experiment by assuming that

---

[89] This last hypothesis (*H3*) predicts no significant difference in the mean extent of lying between both treatments. In order to detect a potential difference here, I will use a cluster bootstrap mean test that accounts for within correlations between repeated observations in different rounds (for more details on this bootstrap procedure, see appendix C.4.). To estimate the statistical power of this bootstrap procedure, I conduct a power analysis based on simulated data with 5,000 repetitions. This power analysis refers to Davidson and MacKinnon (1996). Therefore, I assume that the extent of lying is normally distributed. Here, the mean extent of lying in each round of the baseline treatment is assumed to be the average extent that results from truth-telling (i.e., reporting the optimal investment) and lying to the fullest possible extent (i.e., reporting the maximal investment). Moreover, the standard deviation in each round is assumed to be half of the respective mean. Of course, only advice numbers within the possible range (between 0 and 100) are allowed in the simulation. Finally, I assume a significance level of 5% and a sample size of 30 senders per treatment. Under these assumptions, the statistical power of the bootstrap procedure to detect a difference of 20% in the mean extent of lying between both treatments is 0.942. This provides strong reason to assume that a non-significant result of the bootstrap test would really support *hypothesis H3*.

[90] This is also compatible with basic assumptions of rationality in economic theory.

this extent is reflected in the relationship between the senders' lying behavior and their first-order beliefs about the receivers' (mis)trust.

In general, I expect the extent of lying to increase with the extent of expected mistrust, both with and without an honesty oath. However, there are indications that an honesty oath changes the strategic component of the oath-takers' decision-making process. For instance, Jacquemet, Luchini, Shogren, et al. (2019) suggest that an honesty oath reduces the extent of strategic decision making by fostering commitment to the truth. This is supported by Hergueux et al. (2019), whose findings are consistent with the conjecture that an honesty oath induces less belief-driven strategic thinking. In line with this, I hypothesize that under oath more senders in my experiment will tell the truth regardless of their first-order beliefs.[91] On this basis, I argue that, overall, the senders' decision to tell the truth will depend on whether they expect to be trusted or not (which would result in different truth ratios for senders with different expectations). However, I expect the oath to weaken the impact that the senders' expectations have on their decision to tell the truth (which would reduce the difference in truth ratios between senders with different expectations under oath). This leads to my next hypothesis:

Hypothesis H5. *The honesty oath reduces the relative difference in truth ratios between senders who expect to be trusted and senders who expect to be mistrusted.*

Unfortunately, honesty oaths do not work with everyone (Jacquemet et al., 2020). Hence, there must be some differences between players who follow an honesty oath and players who do not. Jacquemet et al. (2020) suggest that an honesty oath "transforms people with weak preferences for lying into being committed to the truth" (Jacquemet et al., 2020), which implies that the honesty oath encourages a self-selection process that separates people based on their preferences for dishonesty. This raises the question of what the oath does to people who have such strong preferences for lying that they lie even under oath. As argued before, these people are expected to spend more time on solving the inner conflict that arises from being dishonest under oath. It seems unlikely, however, that these particularly dishonest people would lie less strategically because of the oath. On the contrary, I argue that the additional effort to rationalize lying under oath would cause them to deliberate more about *how* to lie, which would most likely increase their extent of strategic decision making. If this conjecture is correct, there should be a testable difference in the belief-behavior relationship of liars between both treatments in my experiment.[92] More precisely, if senders who lie after signing an honesty oath spend more time on deliberating about how to lie, I expect their belief-behavior relationship to be stronger with an honesty oath than without it. From this it follows my last hypothesis:

[91] Note that this would also be consistent with the conjecture that an honesty oath directs the attention of the senders towards moral standards (Mazar et al., 2008), since these standards are independent of the senders' first-order beliefs in the game.

[92] This is under the previously made assumption that the extent of belief-driven strategic deliberation among the senders in my experimental design is reflected in the relationship between their lying behavior and first-order beliefs.

Hypothesis H6. *When considering only liars, the relationship between the extent of lying and the extent of expected mistrust is stronger with an honesty oath (HOT) than without it (BT).*

5. Experimental procedure

The experiment complemented Beck's (2020) experiment at the University of Kassel in February 2018. I recruited a total of 124 undergraduate students in addition to a lecture on game theory. The participation in the experiment was voluntary. According to my post-experimental questionnaire, 56 of my subjects were females and 68 males. Moreover, 57.3% of them were students in economics, 15.3% in engineering, and 14.5% in cultural studies. The remaining 12.9% came from various other fields of study. Since economics students can be expected to assume positions throughout their careers in which business ethics are most relevant (Teixeira & Rocha, 2010), I find it an advantage that most participants of my sample have an economic background.

At the beginning of the experiment, each subject randomly received a sheet of paper with a unique player ID. This ID was used to assign them their role and treatment. Directly afterwards, the senders in the HOT were instructed to take the honesty oath in front of the experimenter by following the procedure described in the design section. The honesty oath that I used is presented in appendix C.2.2. The rest of the experiment was conducted using *classEx* (Giamattei & Lambsdorff, 2019). This is an online tool, which allows subjects to log themselves into the experiment anonymously via their smartphones and make their decisions on screen while sitting in the lab. Firstly, the participants logged themselves into my experiment in *classEx* by entering their player ID. Secondly, they read their instructions for the upcoming games on their screens.[93] Thirdly, they played ten rounds of the CDG as described in the design section. Details on the instructions that I presented to the subjects before the start of the game as well as both players' input screens in the game can be found in appendix C.2.1. Finally, the subjects filled out the post-experimental questionnaire. By completing this step, their active participation in the experiment ended. Up to this point, the experimental procedure lasted about 35 minutes. Before leaving, the subjects received their individual payoffs from different experimenters. The average payoff was 5.73 euros.

6. Results

6.1. Overall effects of the honesty oath

To get a first impression of the effect the honesty oath had on the senders, Figure 4-1 shows the distribution of the extent of lying as an average over all ten rounds for both treatments. It can be seen that the overall average extent of lying (over all rounds and senders) is significantly lower with an honesty oath than without it (BT: 148% vs. HOT: 83%; two-sided Mann-Whitney U test: p = 0.004).

[93] In order to reduce the room for interpretation which behavior of the senders is considered as dishonest in the experiment, lying and truth-telling is called as such in the instructions. For a detailed analysis of the impact of making lying explicit under oath, see Jacquemet, Luchini, Rosaz, et al. (2019).

A similar pattern is apparent when each round is considered separately. Here, the differences in the rounds' average extents of lying between treatments are significant in nine of the ten rounds.[94] These results hold when the comparison is based on pooled data that include observations from all rounds (Bootstrap mean test that accounts for within correlations between rounds[95]: $p = 0.003$). This offers strong support for *hypothesis H2*.

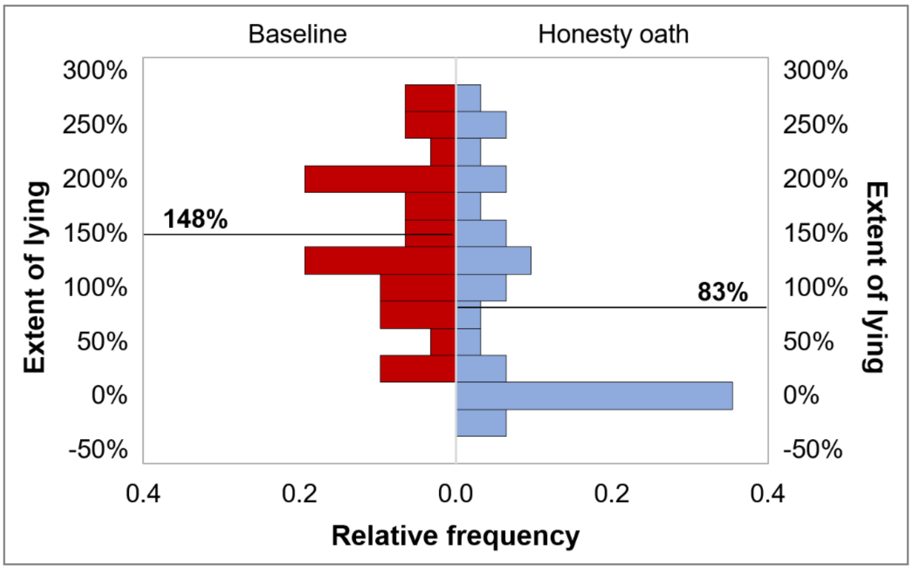


Figure 4-1. Distribution of the extent of lying

Moreover, with an honesty oath, in 51% of observed decisions the senders told the truth, which is an increase of about 132% in comparison to the baseline treatment. This increase in the overall truth ratio (over all rounds and senders) is significant (BT: 22% vs. HOT: 51%; bootstrap unconditional proportion test[96]: $p < 0.001$). This finding is consistent with *hypothesis H1*. To test whether the honesty oath did not only raise the overall truth ratio but also decrease the extent to which *liars* lied, it makes sense to examine liars separately. When only the senders who lied are considered, the average extent of lying is lower with an honesty oath than without it in nine of the ten rounds. However, these differences are not significant in all ten rounds.[97] In addition, a bootstrap mean test on pooled data of all rounds does not detect a significant difference in the extent to which liars lied between both treatments ($p = 0.452$). On this basis, I cannot reject *hypothesis H3*. Beyond that, I do neither detect a significant difference in the distribution nor in the variance of the extent to which liars lied between both treatments.[98] These findings

[94] The results from respective two-sided Mann-Whitney U tests can be found in Table C-1 in appendix C.3.1.
[95] Since all players in my experiment participated in ten rounds of the CDG, the repeated observations between rounds are statistically dependent. To account for this, the bootstrap procedure is implemented by bootstrapping players rather than observations in the sample. For more details, see appendix C.4.
[96] For more details on this bootstrap procedure, see appendix C.4.
[97] The results from respective two-sided Mann-Whitney U tests can be found in Table C-2 in appendix C.3.2.
[98] Kolmogorov-Smirnov and non-parametric Levene's tests that compare the extent to which liars lied between both treatments are not significant in all ten rounds. The results from these tests can be found in Table C-3 in appendix C.3.3.

suggest that subjects who lie under oath barely change their extent of lying due to the oath – an outcome that was not expected by the average player.[99]

Finding 1. *Having subjects sign an honesty oath increases the overall truth ratio but barely reduces the extent to which liars lie* (supporting hypotheses H1 and H3).

6.2. Quick truth-telling and deliberate lying under oath

Turning to the decision time of the senders, there are some similarities between both treatments: Firstly, the overall average decision time (over all rounds and senders) does not differ significantly with and without an honesty oath (BT: 34.4 sec. vs. HOT: 36.2 sec.; two-sided Mann-Whitney U test: $p = 0.550$).[100] This also holds true when the comparison is based on pooled data of all rounds (Bootstrap mean test: $p = 0.462$). Secondly, the decision time per round decreases significantly with the number of rounds played, both with and without an honesty oath (Spearman's rank correlations[101]: BT: $\rho = -0.519$ with $p < 0.001$; HOT: $\rho = -0.535$ with $p < 0.001$). The decrease in decision time over time indicates that the players learned how to decide more efficiently over the course of the experiment.[102]

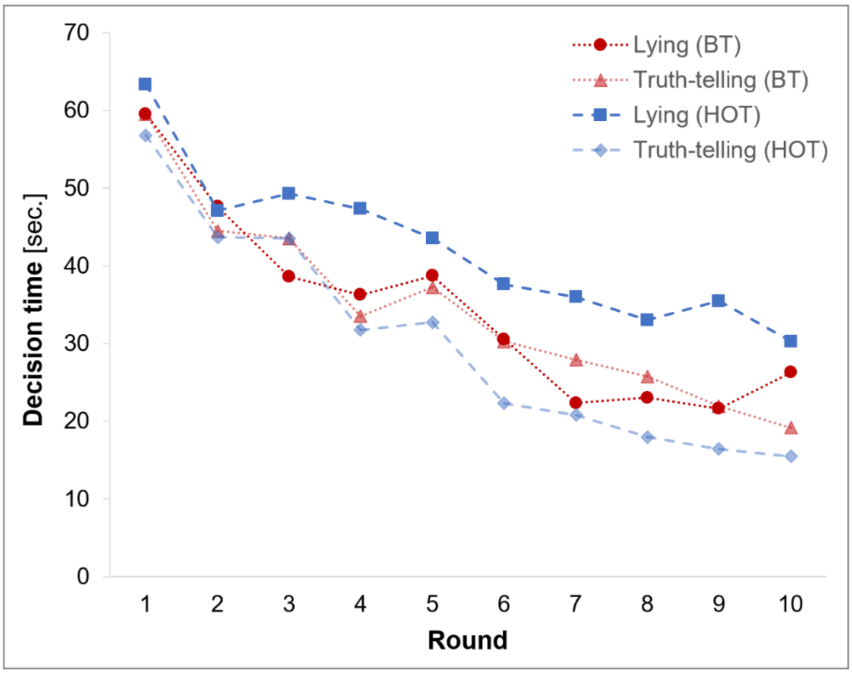


Figure 4-2. Decision time

[99] This can be seen by analyzing how the receivers expected the honesty oath to change the senders' lying behavior: While the receivers (rightly) expected the oath to increase the overall truth ratio among the senders, they (falsely) expected it to reduce the extent to which liars lie. This suggests that the honesty oath disproportionally fostered the receivers' trust in the senders. For visualization and testing of these results, see appendix C.3.7.

[100] For an overview of the average decision time in all rounds and its differences between both treatments, refer to Table C-4 in appendix C.3.4.

[101] Note here that the p-values of all correlations in this paper are based on a bootstrap correlation coefficient test that accounts for within correlations between individual decisions in different rounds. For more details, see appendix C.4.

[102] This is in line with Glöckner (2009) who points out that decision times are expected to decrease when subjects have to repeat a task.

Figure 4-2 visualizes the development of the average decision time over the ten rounds of the experiment with separate graphs for liars and truth-tellers in both treatments. This reveals fundamental differences between both treatments: Without an honesty oath, the average decision time of liars and of truth-tellers does not differ significantly in all ten rounds.[103] This result holds when the comparison is based on pooled data of all rounds (Bootstrap mean test: $p = 0.578$). With an honesty oath, however, it took the senders more time to lie than to tell the truth in all ten rounds. Here, the corresponding differences in the average decision time are significant in most of the rounds.[103] A bootstrap mean test on pooled data of all rounds yields a similar result ($p < 0.001$). The main reason for these results is that in nine of the ten rounds the average decision to lie took longer with an honesty oath than without it. In fact, the average decision time of liars was 34.0 seconds in the baseline treatment and 42.4 seconds in the honesty oath treatment, which is an increase in decision time of about 25%. The corresponding differences in the average decision time are significant in half of all rounds.[103] Performing a bootstrap mean test on pooled data of all rounds confirms this result ($p = 0.005$). This offers strong support for *hypothesis H4a*.

Moreover, on average, truth-telling was faster with an honesty oath than without it in nine of the ten rounds. As a result, the average decision time of truth-tellers was 35.9 seconds in the baseline treatment and 30.1 seconds in the honesty oath treatment, which is a decrease in decision time of about 16%. However, in none of the rounds this difference is significant.[103] This result holds when the comparison is based on pooled data of all rounds (Bootstrap mean test: $p = 0.121$). On this basis, I find only weak support for *hypothesis H4b* – even though the direction of the differences in the decision time of truth-tellers between both treatments is consistent with this hypothesis. These results suggest that the honesty oath intensifies the moral dilemma of liars while (at least slightly) facilitating moral deliberation for truth-tellers.

Finding 2. *The honesty oath increases the decision time of liars* (supporting hypothesis H4a).

[103] The results from respective two-sided Mann-Whitney U tests can be found in Table C-5 in appendix C.3.4.

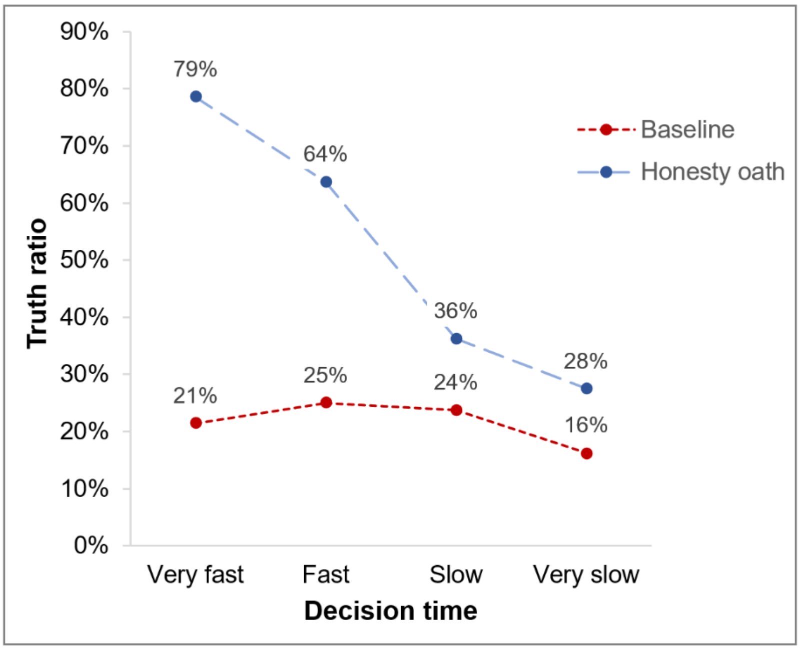


Figure 4-3. Truth ratio and decision time

Figure 4-3 shows the overall truth ratio (over all rounds and senders) as a function of the decision time.[104] The figure illustrates the aforementioned difference in the overall truth ratio between both treatments. However, according to bootstrap unconditional proportion tests, this difference is only significant for "very fast" and "fast" decisions ($p < 0.001$), while it misses significance for "slow" ($p = 0.177$) and "very slow" ($p = 0.234$) decisions. This is mainly because the senders' decision whether to tell the truth or not is related to their decision time only when they signed an honesty oath: Without the oath, the senders do not decide to tell the truth significantly more or less often for different decision times, whereas with the oath, the overall truth ratio differs significantly for different decision times.[105] More precisely, in the honesty oath treatment, the overall truth ratio decreases with increasing decision time.[106] As a result, the chance that a randomly selected sender who has signed an honesty oath told the truth is significantly and about three times higher if the sender decided "very fast" than if they decided "very slowly" (bootstrap unconditional proportion test: $p < 0.001$). These findings illustrate that the honesty

[104] Note that the decision time is displayed in quarters. To take account of the fact that the decision time decreases over the ten rounds of the experiment and that it differs between both treatments, the quarters were determined per round and treatment, and then aggregated over all ten rounds and both treatments. Hence, the four decision time categories indicate how fast (or slow) the respective decisions were in relation to the other decisions that were made under the same conditions.

[105] The results from separate bootstrap unconditional proportion tests that make a pairwise comparison of the different truth ratios between the four decision time categories can be found in Table C-6 (for the baseline treatment) and Table C-7 (for the honesty oath treatment) in appendix C.3.5.

[106] The difference between both treatments can also be shown by analyzing the rank-biserial correlation between the decision time and the dichotomized decision whether to tell the truth or not. In the baseline treatment, this correlation is close to zero and not significant (Spearman's rank correlation: $\rho = 0.028$ with $p = 0.694$). By contrast, in the honesty oath treatment, the same correlation is negative and significant (Spearman's rank correlation: $\rho = -0.353$ with $p < 0.001$). Moreover, the difference between these two correlations is significant (two-sample bootstrap correlation coefficient test: $p < 0.001$; for details on this bootstrap procedure, see appendix C.4.). It can be concluded that the overall truth ratio only decreases with increasing decision time when the senders signed an honesty oath.

oath leads to an increase in the proportion of subjects who engage in quick (and presumably less deliberate) truth-telling.

Finding 3. *The faster subjects decide under oath, the higher is the chance that they tell the truth* (complementing hypothesis H4b).

6.3. Intuitive truth-telling and strategic lying under oath

To visualize the relationship between the lying behavior of the senders and their first-order beliefs, Figure 4-4 displays the overall average extent of lying (over all rounds and senders) for senders who expected to be trusted and for senders who expected to be mistrusted, separately for both treatments. It illustrates – in a simplified form – that the extent of lying increases significantly with the extent of expected mistrust, both with and without an honesty oath (Spearman's rank correlations: BT: $\rho = 0.397$ with $p < 0.001$; HOT: $\rho = 0.422$ with $p < 0.001$). The minor difference between these two correlations is not significant (two-sample bootstrap correlation coefficient test[107]: $p = 0.834$).

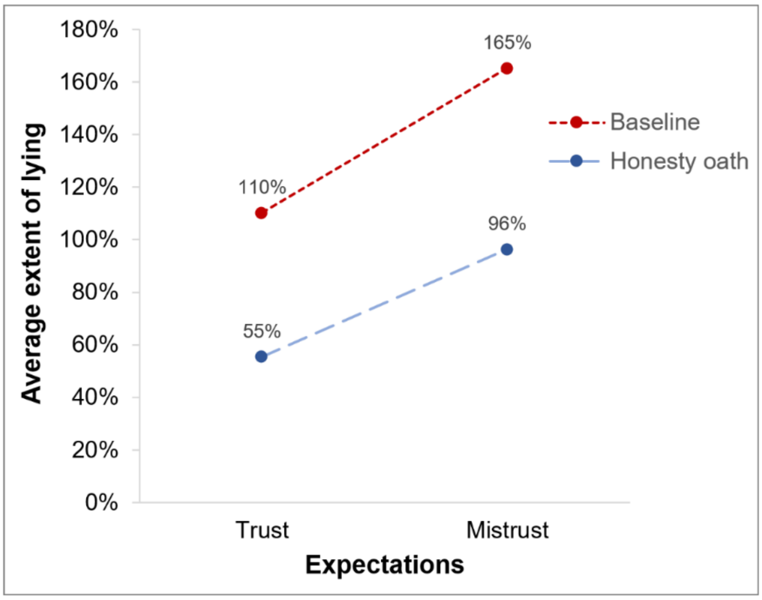
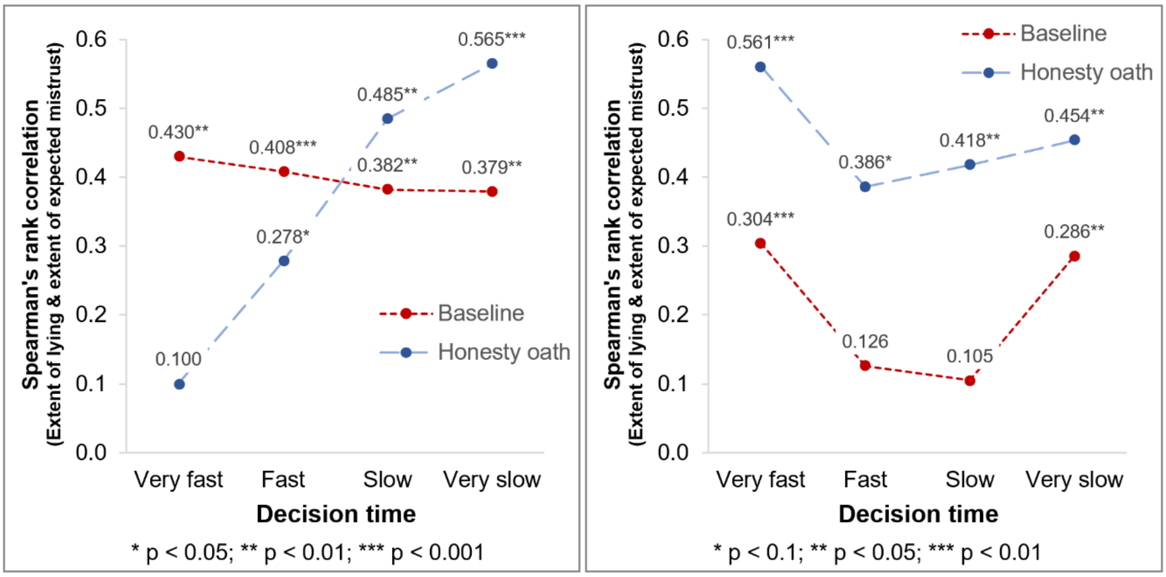


Figure 4-4. Lying behavior and first-order beliefs

Solely on this basis, one might (falsely) conclude that the senders' lying behavior is equally related to their first-order beliefs in both treatments. However, the picture changes when the senders' belief-behavior relationship is examined separately for different decision times.[108] For this purpose, Figure 4-5 shows the correlations between the extent of lying and the extent of expected mistrust for different decision times, separately for both treatments. Note that in the left part of the figure (4-5a) all senders are considered, whereas in the right part (4-5b) only liars are considered.

[107] For more details on this bootstrap procedure, see appendix C.4.
[108] As a side note, the senders' decision time appears not to be related to their first-order beliefs, since their decision time is only weakly and not significantly correlated with their extent of expected mistrust in both treatments (Spearman's rank correlations: BT: $\rho = 0.087$ with $p = 0.239$; HOT: $\rho = 0.054$ with $p = 0.550$).

As can be read from Figure 4-5a, without an honesty oath, the belief-behavior relationship barely changes with the decision time. In fact, a pairwise comparison of the respective correlations in the figure shows that the correlations in the baseline treatment do not differ significantly between different decision times (separate two-sample bootstrap correlation coefficient tests: $p > 0.1$). By contrast, with an honesty oath, the belief-behavior relationship strongly intensifies with the decision time. Here, the correlation coefficient between the extent of lying and the extent of expected mistrust is about six times higher for "very slow" than for "very fast" decisions (Spearman's rank correlations reported in Figure 4-5a). The difference between these two correlations is significant (two-sample bootstrap correlation coefficient test: $p = 0.009$). It must be born in mind, however, that the faster the senders decided, the more of them told the truth. Therefore, this difference might be explained by an unequal distribution of senders who told the truth regardless of their first-order beliefs. To verify this, I will examine the belief-behavior relationship of liars separately.



(a) (b)

Figure 4-5. Relation between lying behavior and first-order beliefs for different decision times: (a) Correlations for all senders; (b) Correlations for only liars

Considering only liars, in Figure 4-5b, it can be seen that their belief-behavior relationship is stronger with an honesty oath than without it for all decision times. As a result, without an honesty oath, the extent to which liars lied and their extent of expected mistrust are significantly positively correlated with each other (Spearman's rank correlation: $\rho = 0.209$ with $p = 0.006$). With an honesty oath, however, the same correlation is stronger and also significant (Spearman's rank correlation: $\rho = 0.449$ with $p < 0.001$). The difference between these two correlations is significant at the 10%-level (two-sample bootstrap correlation coefficient test: $p = 0.086$). This offers weak support for *hypothesis H6*. Firstly, this suggests that the honesty oath intensifies the relationship between the lying behavior and the first-order beliefs of liars. Secondly, these results indicate that the weak belief-behavior relationship of senders who signed an honesty oath and decided "very fast" can indeed be explained by the high fraction of truth-tellers among them.

> **Finding 4.** *The honesty oath strengthens the relationship between the lying behavior and the first-order beliefs of liars* (supporting hypothesis H6).

I will now analyze how the honesty oath influenced the relationship between the senders' decision to tell the truth and their first-order beliefs. In both treatments, I find that senders who expected to be trusted told the truth more often than senders who expected to be mistrusted. However, this difference in the truth ratios between senders with different expectations (i.e., trust expected vs. mistrust expected) is smaller with an honesty oath than without it: To specify, without the oath, the overall truth ratio of senders who expected to be trusted was significantly and about 300% higher than that of senders who expected to be mistrusted (trust expected: 44% vs. mistrust expected: 11%; bootstrap unconditional proportion test: $p < 0.001$). By contrast, with an honesty oath, the overall truth ratio of senders who expected to be trusted was less significantly (i.e., only at the 10%-level) and only about 45% higher than that of senders who expected to be mistrusted (trust expected: 64% vs. mistrust expected: 44%; $p = 0.063$). The comparison of both treatments shows that the oath reduced the relative difference in truth ratios between senders who expected to be trusted and senders who expected to be mistrusted. Here, a bootstrap mean test confirms that the respective truth ratio difference is significantly smaller in the honesty oath treatment than in the baseline treatment (BT: 300% vs. HOT: 45%; $p = 0.004$). These findings are consistent with *hypothesis H5*. This suggests that truth-telling under oath is less related to the senders' first-order beliefs than without the oath.

> **Finding 5.** *Having subjects sign an honesty oath weakens the relationship between their decision to tell the truth and their first-order beliefs* (supporting hypothesis H5).

6.4. Temporal consistency of (dis)honest behavior under oath

To analyze the consistency of the senders' lying behavior over time, it makes sense to differentiate between rounds with similar and identical decision problems. It should be reminded that the investment conditions (i.e., the optimal investments) in two consecutive rounds were never the same. Thus, in two consecutive rounds, the senders were confronted with similar but not identical decision problems. However, each optimal investment value was repeated once with a time lag of five rounds. This ensured that the senders were confronted with identical decision problems with a time lag of five rounds. Based on this, I will use a time lag of one round to determine the temporal consistency of the senders' lying behavior when facing *similar* decision problems and a time lag of five rounds to determine the temporal consistency of their behavior when facing *identical* decision problems.

When confronted with identical decision problems, the lying behavior of the senders was consistent over time, both with and without an honesty oath. This is reflected in the fact that in both treatments the extent of lying is significantly positively correlated with its time-lagged values when the time lag is five rounds (Spearman's rank correlations: BT: $\rho = 0.538$ with $p < 0.001$; HOT: $\rho = 0.691$ with $p < 0.001$). In this case, the difference between these two autocorrelations is not significant (two-sample bootstrap correlation coefficient test: $p = 0.216$). It can be assumed that, for identical decision problems, the honesty oath barely made the senders' lying behavior more consistent over time as their behavior was already highly consistent without the oath.

The picture changes when the senders were confronted with similar but not identical decision problems: Without an honesty oath, the autocorrelation of the extent of lying is significant but weak when the time lag is one round (Spearman's rank correlation: $\rho = 0.175$ with $p = 0.005$). In comparison, with an honesty oath, the same autocorrelation is stronger and also significant (Spearman's rank correlation: $\rho = 0.602$ with $p < 0.001$). Here, the difference between these two autocorrelations is significant (two-sample bootstrap correlation coefficient test: $p < 0.001$). This indicates that the honesty oath increased the temporal consistency of the senders' lying behavior when they were facing similar decision problems.

> **Finding 6.** *Having subjects sign an honesty oath makes their lying behavior under similar conditions more consistent over time.*

Finally, the proportion of senders who never told a lie throughout the entire experiment is significantly higher with an honesty oath than without it (BT: 0% vs. HOT: 29%; two-sided Fisher exact test: $p = 0.002$).[109] This suggests that the oath made the senders tell the truth more consistently over time.

> **Finding 7.** *Having subjects sign an honesty oath makes their decisions to tell the truth more consistent over time.*

6.5. Summary

The honesty oath increased the overall truth ratio but did not significantly reduce the extent to which *liars* lied (*see Finding 1*). In fact, the oath had different effects on how truth-telling and lying occurred: On the one hand, the additional truth-tellers under oath tended to decide rather quick (*see Finding 3*) and, in general, truth-telling under oath was more detached from the decision-makers' first-order beliefs (*see Finding 5*). On the other hand, lying took more time with an honesty oath than without it (*see Finding 2*). In addition, the oath strengthened the relationship between the lying behavior and the first-order beliefs of liars (*see Finding 4*). Finally, the honesty oath made the senders' lying behavior (*see Finding 6*) as well as their decisions to tell the truth (*see Finding 7*) more consistent over time.

[109] Note that this difference between both treatments cannot be explained by the fact that more senders told the truth under oath because when the different truth ratios in both treatments are considered, senders who never lied in all ten rounds are significantly overrepresented *only* in the honesty oath treatment. For a more detailed analysis and testing of these results, see appendix C.3.6.

For an overview of the most important findings, Table 4-1 summarizes the results of my hypotheses testing.

Hypothesis	Prediction	Result
H1	*The overall truth ratio is higher with an honesty oath (HOT) than without it (BT).*	Supported
H2	*The average percentage extent of lying is lower with an honesty oath (HOT) than without it (BT).*	Supported
H3	*The average percentage extent to which liars lie does not differ with an honesty oath (HOT) and without it (BT).*	Not rejected
H4a	*The average decision to lie takes longer with an honesty oath (HOT) than without it (BT).*	Supported
H4b	*The average decision to tell the truth takes shorter with an honesty oath (HOT) than without it (BT).*	Weakly supported
H5	*The honesty oath reduces the relative difference in truth ratios between senders who expect to be trusted and senders who expect to be mistrusted.*	Supported
H6	*When considering only liars, the relationship between the extent of lying and the extent of expected mistrust is stronger with an honesty oath (HOT) than without it (BT).*	Weakly supported

Table 4-1. Hypotheses testing results

In the next section, I will discuss these findings against the backdrop of the existing literature.

7. Discussion

Ethical oaths as a means to foster commitment to moral standards of behavior among business leaders have become more prominent in recent years. In fact, senior executives and employees in some businesses are already obliged to take ethical oaths when coming into office (de Bruin, 2016). The idea behind such oaths is to guide the oath-takers' behavior by aligning it with moral standards (Mazar et al., 2008). In addition, ethical oaths serve to generate trust in the decision makers who have taken the oath (Boatright, 2013). In support of this, I find that the existence of an honesty oath can foster trust in the oath-takers' honesty (*see Finding C-1 in appendix C.3.7.*). Given this and the fact that oaths are frequently used as a management device to enhance truth-telling (Jacquemet, Luchini, Rosaz, et al., 2019), it is a question of utmost practical relevance to better understand how honesty oaths work in terms of inducing the desired behavioral changes. Against this backdrop, in this paper, I analyze causal effects of an obligatory honesty oath on moral deliberation, strategic decision making, and the temporal consistency of (dis)honest behavior.[110] For this purpose, I implement a laboratory experiment using a

[110] In practice, many ethical oaths are not to be taken voluntarily (e.g., the Dutch Banker's Oath or oaths of office for public officials). For this reason, the oath in my experiment is made a precondition for participating in the honesty oath treatment. It bears repeating that I do not argue in favor of obligatory over voluntary oath-taking (for a detailed discussion of this issue, see Jacquemet et al., 2013). However, since obligatory oaths are typical for both the private and the public sector, my study aims at examining the effects of an obligatory honesty oath. For more details on the oath procedure that I used, see section 3.2.

between-subjects design with repeated measurements. I use Beck's (2020) Continuous Deception Game (CDG), which introduces continuous variables to the sender-receiver game.[111]

Consistent with previous studies (e.g., Beck et al., 2020; Carlsson et al., 2013; Jacquemet et al., 2013; Jacquemet et al., 2018; Kemper et al., 2016), I provide further evidence that an honesty oath leads to more truth-telling. Beyond that, I find that the oath has no significant effect on the extent to which liars lie (*see Finding 1*), which indicates that people who take an honesty oath either entirely follow the oath or totally disregard it. This interpretation of my results is in line with Jacquemet et al. (2020) who observe a similar effect of an honesty oath in an income/tax declaration experiment. However, Jacquemet et al. (2020) also find that the oath affects only partial liars. In this aspect, my results differ from those of Jacquemet et al. (2020), as my findings suggest that whether people follow an honesty oath or not does not depend on the potential sizes of their lies.

To explain the difference between my results and those of Jacquemet et al. (2020), it makes sense to compare the decision problems, which players face in both experiments. Both experiments have in common that the players are confronted with a similar trade-off dilemma between financial gain from cheating and the moral costs of lying (Mazar et al., 2008). Therefore, subjects in both experiments make their decisions based on their individual preferences for lying for financial gain. However, decision making requires more strategic reasoning in my experiment than in the one of Jacquemet et al. (2020). The reason for this is that a lie in Jacquemet et al.'s (2020) income declaration task does not have to be convincing in order to yield a favorable result for the liar. This means that there is no need for the subjects to be concerned about their potential to manipulate another person when lying. By contrast, the senders in my experiment have to consider to what extent the receivers will mistrust their lies (for a detailed discussion of considerations about the potential to manipulate others in the CDG, see Beck, 2020). Therefore, it is possible that senders who tell partial lies under oath are still inclined to lie to the highest extent they expect to be convincing. This difference between both experiments is able to explain why the honesty oath in my experiment affects liars with all sizes of lies, while it affects only partial liars in the study of Jacquemet et al. (2020). Moreover, the additional strategic component in my experiment highlights the important role that strategic considerations about the behavior of others can play under oath.

This raises the question of how an honesty oath impacts strategic considerations of oath-takers when they decide about lying? To answer this question, I refer to previous studies that provide evidence that strategic considerations when lying are reflected in the relationship between the displayed lying behavior and beliefs about others (e.g., López-Pérez & Spiegelman, 2013; Lundquist et al., 2009; Peeters et al.,

[111] What many other studies that examine the honesty-fostering effects of oaths have in common is that they observe lying behavior as a discrete choice (e.g., the binary choice between truth-telling and lying). I use Beck's (2020) CDG because it allows for observation of more differentiated lying behavior (i.e., the individual extent of lying) and first-order beliefs (i.e., the individual extent of expected mistrust). This enables me to better understand how lying is affected by an honesty oath in strategic interaction.

2015). In the light of this, I find that the honesty oath strengthens the relationship between the lying behavior and the first-order beliefs of liars (*see Finding 4*), while it weakens the relationship between truth-telling and the decision-makers' first-order beliefs (*see Finding 5*). This indicates that the honesty oath has fundamentally different effects on how truth-telling and lying occur. I will address these effects separately and in more detail now.

To begin with, I will discuss how truth-telling occurs under oath. In line with Peeters et al. (2015), I provide evidence that whether subjects decide to tell the truth or not in the absence of an oath is closely linked to whether they expect to be trusted or not. When subjects take an honesty oath however, this link between truth-telling and the subjects' first-order beliefs becomes much weaker (*see Finding 5*). This suggests that under oath people tend to tell the truth regardless of their beliefs about others. These findings are in line with Hergueux et al. (2019) who show that an honesty oath reduces the extent of strategic thinking in the public goods game. Beyond that, my results complement those of Jacquemet, Luchini, Shogren, et al. (2019) who examine the effects of an honesty oath and two other oaths that aim at fostering cognitive effort in a discrete choice experiment. By comparing the effects of these oaths, Jacquemet, Luchini, Shogren, et al. (2019) suggest that an honesty oath does not foster cognitive reasoning but the effort to tell the truth. On this basis, they argue that an honesty oath reduces the tendency to be strategic. It should be noted, however, that Jacquemet, Luchini, Shogren, et al.'s (2019) discrete choice experiment does not feature strategic interaction. Therefore, strategic behavior in their experiment refers to the intentional misrepresentation of choice preferences. By contrast, strategic lying in the CDG requires building first-order beliefs about the other player's mistrusting behavior (Beck, 2020). With that in mind, my findings add to those of Jacquemet, Luchini, Shogren, et al. (2019) that an honesty oath leads to less belief-driven and more intuitive truth-telling when the decision is made in a strategic environment in which the decision maker's financial success depends on the behavior of another person.

In support of this, I find that, the faster subjects decide under oath, the higher is the chance that they tell the truth (*see Finding 3*). Under the assumption that the extent of cognitive reasoning is reflected in the decision time (e.g., Hergueux et al., 2019; Jacquemet, Luchini, Rosaz, et al., 2019; Kahneman, 2003, 2011; Rubinstein, 2007; Schunk & Betsch, 2006), this suggests that, after taking an honesty oath, truth-tellers deliberate less about their decisions than liars. This in turn is another indication that truth-telling is more intuitive under oath (Kahneman, 2003, 2011).

Unfortunately, honesty oaths do not work with everyone (Jacquemet et al., 2020; Jacquemet, Luchini, Rosaz, et al., 2019). Therefore, I will now discuss how lying occurs under oath. In support of previous studies, I provide further evidence that an honesty oath increases the decision time of liars (*see Finding 2*), which suggests that the oath causes liars to deliberate more about their lies (Kahneman, 2003, 2011; Rubinstein, 2007; Schunk & Betsch, 2006). This is consistent with Jacquemet, Luchini, Rosaz, et al. (2019) who show that an honesty oath can increase the decision time of liars in Erat and Gneezy's (2012) sender-receiver game.

But what happens in the additional time that liars take to decide under oath? One explanation is that the honesty oath intensifies the moral dilemma of liars, which causes them to deliberate more on whether to lie or not. Another explanation is that liars under oath spend more time to reason about how to lie. Based on changes in subjects' decision times, Jacquemet, Luchini, Rosaz, et al. (2019) argue in favor of the first explanation. However, my results also offer support for the second explanation. To be more precise, I find that the honesty oath strengthens the relationship between the extent to which liars lie and the extent to which they expect to be mistrusted (*see Finding 4*). Under the assumption that strategic considerations when lying are reflected in this belief-behavior relationship (e.g., López-Pérez & Spiegelman, 2013; Lundquist et al., 2009; Peeters et al., 2015), this suggests that liars decide more strategically under oath. Note that this finding is not contradicting Jacquemet, Luchini, Rosaz, et al. (2019) because it also allows for the interpretation that the honesty oath intensifies strategic reasoning and moral deliberation of liars at the same time.

At first glance, however, my results seem to stand in contrast to Hergueux et al. (2019) who argue that subjects who take an honesty oath put less weight on strategic thinking. Hergueux et al. (2019) base this conclusion on the observation that an honesty oath strengthens the relationship between estimated and actual contributions in the public goods game. What Hergueux et al.'s (2019) and my study have in common is that the subjects decide in an environment in which the behavior of others has an impact on their payoffs. In addition, in both of our studies, the honesty oath strengthens the relationship between the observed behavior and stated beliefs about the behavior of others. However, the obvious differences in our experimental designs (public goods game vs. sender-receiver game) demand for diverging interpretations of our results: In Hergueux et al.'s (2019) experiment a stronger belief-behavior relationship indicates more reciprocal behavior, whereas in my experiment it suggests that players deliberate more about how their lies can manipulate others. This comparison provides reason to assume that an honesty oath makes oath-takers consider the effects their behavior has on others more carefully; and whether this induces strategic thinking or not depends on the context in which the oath is taken.

Finally, I provide evidence that having subjects sign an honesty oath makes both their lying and truth-telling behavior more consistent over time (*see Findings 6 and 7*). Again, this speaks in favor of the conjecture that the honesty oath directs the oath-takers attention towards moral standards (Mazar et al., 2008) and confronts them with the question of who they want to be in the experiment (Blok, 2013). This suggests that an honesty oath aligns the oath-takers' behavior to some internal reference point of what is right or wrong in a given context. In support of this, Hergueux et al. (2019) find that individuals behave more consistently with their internalized social norms after taking an honesty oath. Following this line of argument, my findings indicate that the honesty oath encourages the subjects to make one basic decision whether to commit to truth-telling throughout the entire experiment or to disregard the oath and accept being dishonest in order to enrich themselves. Thus, if subjects decide to follow the oath, there is no need for them to deliberate more about their decisions in the game or to make strategic considerations when telling the truth. If they decide not to follow the oath however, the oath intensifies their moral dilemma and makes them consider how lying manipulates the other player more carefully.

Against the backdrop that obligatory oaths are frequently used as a management device to foster honest behavior (Blok, 2013; Boatright, 2013; de Bruin, 2016), my findings offer several business implications. On the one hand, when people are given the chance to behave opportunistically, oath-taking can enhance quick and intuitive truth-telling, at least when the oath is taken shortly before the decision in question is made. This is because an honesty oath can reduce the amount of strategic reasoning when deciding whether to tell the truth or to lie for personal financial gain. Thus, honesty oaths can cause oath-takers to tell the truth more consistently over time. On the other hand, when people lie under oath, the oath barely changes the sizes of their lies. In addition, oath-taking can make dishonest behavior more deliberate and strategic. This poses a potential risk because it demonstrates that particularly dishonest people, who lie even under oath, might deliberate more about how they can manipulate others when lying *due* to the oath. Depending on the given context, this could lead to better-thought-out lies and, thus, to more effective opportunistic behavior – especially since honesty oaths foster trust in the oath-takers. As a result, an honesty oath can make some individuals lie more consistently over time. While such potential negative side effects of oath-taking suggest an interesting field for future research, my findings strongly imply that the positive effects of honesty oaths outweigh such potential risks.

CHAPTER 5

Conclusion and future work

This dissertation aims at investigating whether and how distinct measures that are supposed to foster honest behavior work. It consists of three papers, in which I analyze the effectiveness and mechanisms of such measures on the basis of several laboratory experiments. In the first paper, I find that honesty oaths perform better in terms of lie mitigation compared to the four-eyes principle and monitoring (*see Chapter 2*). In the second paper, I design a new experiment that allows the measurement of more differentiated lying behavior in a two-player relationship of strategic interaction (*see Chapter 3*). By means of this new experimental design, I find, in the third paper, that an honesty oath can enhance quick and intuitive truth-telling but also lead to more deliberate and strategic lying (*see Chapter 4*).

Taken together, these three papers improve the understanding of why intrinsic motivators, such as an honesty oath, can be much more effective in fostering people's honesty than controlling measures, such as the four-eyes principle or monitoring. The most important reason for this is that an honesty oath appeals directly to the oath-takers' honesty. This confronts them not only with the question of what to do in the context addressed by the oath but also with the question of who they want to be in that context (Blok, 2013). Thus, an honesty oath does not induce morality in otherwise immoral people. Instead, it reminds the oath-takers of their moral standards (Mazar et al., 2008), which must have been already internalized before taking the oath. In other words, an honesty oath does not create intrinsic motivation to behave honestly out of nowhere, but it fosters pre-existing intrinsic motivation by directing the oath-takers' attention to some internal reference point of what is right or wrong in a given context (Mazar et al., 2008). This is exactly what most controlling measures fail to do. By contrast, they direct people's attention to an external reference point of what is allowed and what is forbidden. Thereby, such measures impose external control on individuals in order to force them to comply with rules and moral standards. Depending on the context, this can heavily undermine honest behavior because external control can induce the feeling of being mistrusted, which in turn can crowd out the intrinsic motivation to behave honestly (Banfield, 1975; Fehr & List, 2004; Kreps, 1997; Schulze & Frank, 2003).

This does not imply that regulatory authorities should completely abandon measures that create extrinsic incentives to comply with ethical standards, since such measures might also provide more transparency and accountability, which are also highly important for fighting fraud, corruption and other unethical business conduct (Gaventa & McGee, 2013; OECD, 2007; UNIDO, 2020). Nevertheless, my findings provide strong reason to question whether such controlling measures alone are sufficient to prevent people from behaving dishonestly. Hence, instead of relying on measures that try to force ethical standards upon people, firms and regulatory authorities should focus more on creating an environment that intrinsically motivates people to be good persons.

Strength and limitations

As the empirical findings in my dissertation are based on observations made in laboratory experiments, they lack the external validity of field studies. However, the controlled environment in the lab allows me to manipulate individual factors in the decision-making process and isolate their effects on people's (dis)honesty.[112] Against this backdrop, it would be highly interesting to test whether my key findings can be replicated in the field. For instance, I suggest future field experiments to verify the effectiveness of oath-taking in the workplace or at least in real work tasks.[113] Moreover, I suggest to verify my findings on the effects of an honesty oath on players' decision time by analyzing how oath-taking affects the time that people spend on real work tasks that provide an opportunity to engage in dishonest behavior.

Furthermore, it should be noted that the majority of my participants are undergraduate students with an economic background. For this reason, my findings are not representative for the general population of people who are responsible for making decisions that are relevant from an ethical perspective. However, economics and business students can be "expected to be our future business people and potentially the economic leaders and politicians of tomorrow" (Teixeira & Rocha, 2010, p. 663). Therefore, they are likely to assume positions in which business ethics are most relevant. With that in mind, analyzing a sample that is biased by students with an economic background might not be a disadvantage when the analysis focuses on measures that are inspired by business ethics. Nevertheless, when interpreting my findings, it is important to be aware of the origin of my sample.

Future research

Apart from the above suggested field studies, my findings provide a broad basis for future research. To begin with, in the first paper (*see Chapter 2*), the honesty oath is the only measure against dishonest behavior that aims at fostering intrinsic motivation to tell the truth. This raises the question of how other

[112] For a detailed analysis of advantages and disadvantages of examining dishonest behavior in the lab, see Armantier and Boly (2008).

[113] This could be done, for example, by introducing an honesty oath to the field experiment of Cagala et al. (2014), in which the outcome variable (i.e., people's dishonesty) is studied by the likelihood of students stealing a high-quality pen after an exam. To be more specific, Cagala et al.'s (2014) field study could be modified by asking half of the students to voluntarily sign an honesty oath prior to their exam. On this basis, the effectiveness of the oath could be tested by analyzing the impact of the oath on the likelihood that the students steal the pen. Note that, as shown by Jacquemet, Luchini, Rosaz, et al. (2019), most students can be expected to freely sign an honesty oath when asked whether they would like to take it or not. Also, the oath should be formulated in such a way that it covers the stealing of the pen (e.g., by covering the students' actions not only during but also in the context of the exam). Moreover, it is important to ensure that conducting this study does not affect the students' performance in the exam. For this reason, the students should not feel pressured to take the oath.

comparable measures, such as business codes (Kaptein, 2004) or ethical leadership (Brown & Treviño, 2006), would perform compared to the honesty oath.

In addition, in the first and the third paper, I observe only short-term effects of oath-taking. Thus, it would be interesting to test how taking an honesty oath affects the oath-takers' honesty in the long run (Pe'er & Feldman, 2020). Moreover, to the best of my knowledge, the third paper (*see Chapter 4*) is the first to address potential negative side effects of oath-taking regarding an increase in strategic decision making of liars under oath. While I do not believe (and find no evidence) that an honesty oath does more harm than good, this also suggests an intriguing field for future research.

Another interesting result, which is barely in the focus of my analysis, is that the honesty oath in the third paper disproportionally fosters the trust that other players have in the oath-takers. This increase in honesty-related trust is another side effect of oath-taking that has been barely addressed in previous empirical studies (one of the few studies that touch this issue is Jacquemet et al., 2018). Since trust is an important aspect of strategic interaction (Schumacher, 2006), it would be interesting to analyze the trust-fostering effects of the honesty oath in more detail.

Furthermore, in the process of developing the CDG in the second paper (*see Chapter 3*), I designed several similar but slightly different versions of this experiment. The final version is the one that found most approval in colloquia and fitted best to my research question in the third paper. However, the other (unpublished and partially unfinished) versions of the CDG made me realize the potential to modify this experimental design in order to examine different aspects of lying and mistrusting behavior. One example can be taken from the second paper: Here, my results suggest that sophisticated liars take their expected credibility in front of others into account when they decide about lying. In particular, I find that liars (rightly) believe that their lies will be disproportionally mistrusted if they lie to an unrealistically high extent. This finding is primarily based on players' first-order beliefs, since I decided not to observe higher-order beliefs.[114] However, there could be additional reasons why people do not lie to the fullest possible extent. For instance, potential liars could be concerned about their social identity and, therefore, try to avoid being perceived as dishonest (as suggested by Gneezy et al., 2018). To distinguish such social identity concerns from strategic considerations of liars about their potential to manipulate others when lying, it would be interesting to elicit not only first- but also second-order beliefs in the CDG.

Finally, in this dissertation, I observe the isolated effects of individual measures against dishonesty. Both in the public and in the private sector, it is common practice, however, to implement a mix of several of such measures. Maybe the four-eyes principle performs better in combination with other measures that aim at fostering the intrinsic motivation to behave honestly? Or, maybe the positive effects

[114] The main reason why I asked the players in my experiment solely for their first-order beliefs is that there is evidence that first-order beliefs already sufficiently capture the relation between beliefs and behavior in sender-receiver games, while second-order beliefs do not provide much more insight (López-Pérez & Spiegelman, 2013).

of oath-taking can be undermined by monitoring the oath-takers? To answer these questions, I suggest analyzing combinations of such measures in suitable laboratory experiments, such as the CDG.

As this wide range of suggestions for future research is based on new findings presented in this dissertation, it is testimony to how this dissertation itself improves the understanding of whether and how different measures against dishonest behavior work. But, more importantly, these suggestions provide direction for further work with the aim of understanding the complex mechanisms of (dis)honest behavior – since, in the end, this dissertation (as all individual research activities in this area) plays only a small part in the bigger picture of enhancing ethical standards in the business world.

Appendices

This section contains the appendices to Chapter 2 (*in subsection A.*), Chapter 3 (*in subsection B.*), and Chapter 4 (*in subsection C.*).

A. Appendix to Chapter 2: Factor approximation in our theoretical model of the utility of lying

In the results section of the paper in Chapter 2, we find evidence in support of some of the major predictions we made for behavioral differences between treatments based on our theoretical model of the utility of lying. Ex post, we can use this model for a more detailed interpretation of our findings. Thus, we estimate the most essential key factors of the model by using the observed mean values of payoffs per treatment.

In each treatment, the mean value of payoffs under complete honesty ($\bar{r}$) is expected to approximate €2.50. Hence, we assume: $\bar{r} = €2.50$ for each treatment.

A.1. Baseline treatment

The observed mean value of payoffs in the baseline treatment is $\bar{p}_{baseline} = €3.67$. Assuming that on average all players engaged in the optimal amount of lying, we get:

$$\left(\frac{1}{2\bar{\delta}_{baseline}}\right)^2 - \bar{r} = \bar{p}_{baseline} - \bar{r} \tag{15}$$

with $\bar{\delta}_{baseline}$ being the average dislike of lying for oneself in the baseline treatment.

Solving this equation for $\bar{\delta}_{baseline}$, we get:

$$\bar{\delta}_{baseline} = \frac{1}{2\sqrt{\bar{p}_{baseline}}} = 0.26\,. \tag{16}$$

A.2. Moral awareness treatment

The observed mean value of payoffs in our moral awareness treatment is $\bar{p}_{moral} = €2.66$. Again, assuming that on average all players engaged in the optimal amount of lying, we get:

$$\left(\frac{1}{2\bar{\alpha}_{moral}\bar{\delta}_{moral}}\right)^2 - \bar{r} = \bar{p}_{moral} - \bar{r} \tag{17}$$

with $\bar{\delta}_{moral}$ being the average dislike of lying for oneself, and $\bar{\alpha}_{moral}$ being the factor to which the honesty oath increased moral awareness on average compared to the baseline treatment.

Solving this equation for $\bar{\alpha}_{moral}$, we get:

$$\bar{\alpha}_{moral} = \frac{1}{2\bar{\delta}_{moral}\sqrt{\bar{p}_{moral}}} . \quad (18)$$

Since we already capture the difference in moral awareness between the baseline and moral awareness treatment by using the factor $\bar{\alpha}_{moral}$, we assume that the average dislike of lying for oneself is equal in these two treatments ($\bar{\delta}_{moral} = \bar{\delta}_{baseline}$).

On this basis, we can calculate $\bar{\alpha}_{moral}$:

$$\bar{\alpha}_{moral} = \frac{1}{2\bar{\delta}_{baseline}\sqrt{\bar{p}_{moral}}} = 1.17 . \quad (19)$$

The interpretation of this factor is that the honesty oath increased moral awareness on average by about 17%.

A.3. Monitoring treatments

The observed mean value of payoffs in the pooled monitoring treatments is $\bar{p}_{monitor} = €3.78$. Again, assuming that on average all players engaged in the optimal amount of lying, we get:

$$\left(\frac{1}{\bar{\alpha}_{monitor}\bar{\delta}_{monitor}}\right)^2 - \bar{r} = \bar{p}_{monitor} - \bar{r} \quad (20)$$

with $\bar{\delta}_{monitor}$ being the average dislike of lying for oneself, and $\bar{\alpha}_{monitor}$ being the factor to which monitoring increased moral awareness on average compared to the baseline treatment.

Solving this equation for $\bar{\delta}_{monitor}$, we get:

$$\bar{\delta}_{monitor} = \frac{1}{\bar{\alpha}_{monitor}\sqrt{\bar{p}_{monitor}}} . \quad (21)$$

Under the assumption that $1 \leq \bar{\alpha}_{monitor} \leq \bar{\alpha}_{moral}$, we can estimate $\bar{\delta}_{monitor}$:

$$0.44 = \frac{1}{\bar{\alpha}_{moral}\sqrt{\bar{p}_{monitor}}} \leq \bar{\delta}_{monitor} \leq \frac{1}{\sqrt{\bar{p}_{monitor}}} = 0.51 . \quad (22)$$

Comparing this to the baseline treatment, we get:

$$1.69 * \bar{\delta}_{baseline} \leq \bar{\delta}_{monitor} \leq 1.98 * \bar{\delta}_{baseline} . \quad (23)$$

This means that the average dislike of lying for oneself increased by at least 69% due to the presence of the monitor (while already considering the potential effects of an increase in players' moral awareness on their honesty due to monitoring). These positive effects of monitoring, however, are overshadowed by the division of moral costs between both players ($n = 2$), since:

$$\frac{\bar{\alpha}_{monitor}\bar{\delta}_{monitor}}{n} = \frac{1.98 * \bar{\delta}_{baseline}}{2} < \bar{\delta}_{baseline} . \tag{24}$$

A.4. Reciprocity treatments

As we find no significant difference between both reciprocity treatments, we define $\bar{p}_{reci} = €2.89$ as the observed mean value of payoffs in the pooled reciprocity treatments.

Again, assuming that on average all players engaged in the optimal amount of lying, we get:

$$\left(\frac{\bar{\beta}_{reci}}{2\bar{\delta}_{reci}}\right)^2 - \bar{r} = \bar{p}_{reci} - \bar{r} \tag{25}$$

with $\bar{\delta}_{reci}$ being the average dislike of lying for another person, and $\bar{\beta}_{reci}$ being the factor to which players care about others on average.

Solving this equation for $\bar{\beta}_{reci}$, we get:

$$\bar{\beta}_{reci} = 2\bar{\delta}_{reci}\sqrt{\bar{p}_{reci}} . \tag{26}$$

A.5. Group treatment

The observed mean value of payoffs in the group treatments is $\bar{p}_{group} = €4.14$. Again, assuming that on average all players engaged in the optimal amount of lying, we get:

$$\left(\frac{1 + \bar{\beta}_{group}}{\bar{\delta}_{group;i} + \bar{\delta}_{group;j}}\right)^2 - \bar{r} = \bar{p}_{group} - \bar{r} \tag{27}$$

with $\bar{\delta}_{group;i}$ being the average dislike of lying for oneself, $\bar{\delta}_{group;j}$ being the average dislike of lying for another person, and $\bar{\beta}_{group}$ being the factor to which players care about others on average.

Solving this equation for $\bar{\beta}_{group}$, we get:

$$\bar{\beta}_{group} = \left(\bar{\delta}_{group;i} + \bar{\delta}_{group;j}\right)\sqrt{\bar{p}_{group}} - 1 . \tag{28}$$

Furthermore, we can assume that both the factor to which players care about others and their dislike of lying for another person do not change between treatments ($\bar{\beta} := \bar{\beta}_{reci} = \bar{\beta}_{group}$ and $\bar{\delta}_j := \bar{\delta}_{reci} = \bar{\delta}_{group;j}$), since we already consider the number of other players affected by lying (m) and the number of players participating in the decision (n) separately.

This implies that we can equate (26) with (28):

$$2\bar{\delta}_j\sqrt{\bar{p}_{reci}} = \left(\bar{\delta}_{group;i} + \bar{\delta}_j\right)\sqrt{\bar{p}_{group}} - 1\,. \tag{29}$$

Since we argue that the dislike of lying for oneself is higher than the dislike of lying for another person ($\bar{\delta}_{group;i} > \bar{\delta}_j$), we get:

$$2\bar{\delta}_j\sqrt{\bar{p}_{reci}} < 2\bar{\delta}_{group;i}\sqrt{\bar{p}_{group}} - 1\,. \tag{30}$$

Solving this inequation for $\bar{\delta}_j$ yields:

$$\bar{\delta}_j < \bar{\delta}_{group;i} * \sqrt{\frac{\bar{p}_{group}}{\bar{p}_{reci}}} - \frac{1}{2\sqrt{\bar{p}_{reci}}}\,. \tag{31}$$

Here we have to estimate $\bar{\delta}_{group;i}$. If we approximate the dislike of lying for oneself in the group treatment with the corresponding value from the baseline treatment ($\bar{\delta}_{group;i} \approx \bar{\delta}_{baseline}$), we get:

$$\bar{\delta}_j < \bar{\delta}_{baseline} * \sqrt{\frac{\bar{p}_{group}}{\bar{p}_{reci}}} - \frac{1}{2\sqrt{\bar{p}_{reci}}} = 0.26 * \sqrt{\frac{\bar{p}_{group}}{\bar{p}_{reci}}} - \frac{1}{2\sqrt{\bar{p}_{reci}}} = 0.017\,. \tag{32}$$

If that approximation is somewhat correct, this inequation has a very intuitive interpretation: The dislike of lying for another person in our treatments was extraordinarily low (meaning that the dislike of lying for oneself in the baseline treatment was at least 15 times higher than the dislike of lying for another person in the respective treatments).

Moreover, this indicates that the average factor to which players care about others ($\bar{\beta}$) was also low, since:

$$\bar{\beta} = \bar{\beta}_{reci} = 2\bar{\delta}_{reci}\sqrt{\bar{p}_{reci}} = 2\bar{\delta}_j\sqrt{\bar{p}_{reci}} = 3.4 * \bar{\delta}_j < 3.4 * 0.017 = 0.058\,. \tag{33}$$

A.6. Summary

According to our theoretical model of the utility of lying, we can interpret our findings in more detail:

(1) The honesty oath increased the moral awareness of our participants by about 17% (≈20%) on average.
(2) The dislike of lying for oneself was at least 15 times higher than the dislike of lying for another person.
(3) This was due to the fact that the dislike of lying for another person was weak.
(4) The degree to which players cared about others in the experiment was low.

(5) Monitoring increased the dislike of lying for oneself by at least 69% (≈70%) (already taking into account the potential effects of an increase in players' moral awareness on their honesty due to monitoring). However, the potential positive effects of monitoring were still overshadowed by the division of moral costs between both players.

B. Appendices to Chapter 3

This section consists of five separate appendices to Chapter 3. The first appendix shows the derivation of the set of game theoretical equilibria of the Continuous Deception Game (CDG) and discusses how rational players would adapt their behavior to their beliefs in equilibrium (*B.1.*). The second appendix contains the instructions and input screens that I presented to the subjects in the CDG (*B.2.*). The third appendix presents an analysis of the consistency of my interpretation of both players' strategies in the CDG with their ex post self-evaluation of their behavior within the experiment (*B.3.*). The fourth appendix analyzes the temporal consistency of both players' behavior and first-order beliefs (*B.4.*). Lastly, the fifth appendix provides details on the bootstrap procedures that I use in the paper in Chapter 3 (*B.5.*).

B.1. Set of game theoretical equilibria

In this appendix, I solve the CDG by identifying its set of game theoretical equilibria, which allows me to determine strategies that are more likely to be pursued by rational players. Note here that the CDG is a sequential game with incomplete information.[115] For the analysis, both players are modeled as risk-neutral rational players who seek to maximize their expected utility based on their beliefs about the other player. In a first step, I derive the set of game theoretical equilibria with only monetary motivation (*B.1.1.*). In a second step, I derive the set of game theoretical equilibria with monetary and non-monetary motivation (*B.1.2.*).

B.1.1. Only monetary motivation

For the beginning, suppose, for simplicity, that both players are *homines oeconomici* who solely value their monetary payoffs and, therefore, do not care about other factors, such as being (or being recognized as) honest or trusting.

I start by analyzing the investor's payoff structure: Given a predefined maximal investment $i_{max} > 0$ and an optimal investment $i^* \in [0, i_{max}]$, which is randomly determined by a uniform distribution ($P([0, i^*]) = \frac{i^*}{i_{max}}$ with $i^* \in [0, i_{max}]$), the investor's payoff $\pi_I(i)$ is defined as the following function of the investment $i \in [0, i_{max}]$:

$$\pi_I(i) = \begin{cases} i \leq i^*: & i_{max} + m_I * i \\ i > i^*: & i_{max} + m_I * (2 * i^* - i) \end{cases} \tag{34}$$

with $m_I > 0$.

[115] If nature is assumed to be another player *and* if only monetary incentives are considered, this game can be thought of as a game with imperfect information where nature makes the first move by choosing the optimal investment i^*, but the investor does not observe nature's move (for more details, see Harsanyi, 1967, 1968a, 1968b).

Note here that the optimal investment i^*, which is not known to the investor, maximizes the investor's payoff function $\pi_I(i)$ by design. Since the *homo oeconomicus* type of investor has no reason to trust the advice a, he or she disregards this information and, therefore, assumes that the true optimal investment i^* is located somewhere between 0 and the maximal investment i_{max} with equal probability. On this basis, the *investor's expected payoff* $\pi^e_{I;hoec}(i)$ can be defined as the following function of the investment i:

$$\pi^e_{I;hoec}(i) = \frac{1}{i_{max}} \int_0^{i_{max}} \pi_I(i)\, di^*. \tag{35}$$

Notice that the domain of the investor's payoff function $\pi_I(i)$ is split into two regions that depend on the location of the optimal investment i^*. For that reason, it is necessary to distinguish between one region in which the optimal investment i^* is lower than the investment i ($i^* < i$) and another in which it is equal or superior to the investment i ($i^* \geq i$). This yields:

$$\begin{aligned}
\pi^e_{I;hoec}(i) &= \frac{1}{i_{max}} \int_0^{i_{max}} \pi_I(i)\, di^* \\
&= \frac{1}{i_{max}} * \lim_{b \nearrow i} \left(\int_0^b \underbrace{\pi_I(b)}_{i^*<i}\, di^* \right) + \frac{1}{i_{max}} \int_i^{i_{max}} \underbrace{\pi_I(i)}_{i^* \geq i}\, di^* \\
&= \frac{1}{i_{max}} * \lim_{b \nearrow i} \left(\int_0^b \big(i_{max} + m_I * (2 * i^* - b)\big)\, di^* \right) + \frac{1}{i_{max}} \int_i^{i_{max}} (i_{max} + m_I * i)\, di^* \\
&= \frac{1}{i_{max}} \int_0^i \big(i_{max} + m_I * (2 * i^* - i)\big)\, di^* + \frac{1}{i_{max}} \int_i^{i_{max}} (i_{max} + m_I * i)\, di^* \\
&= \frac{1}{i_{max}} * \big[i_{max} * i^* + m_I * \big(i^{*2} - i * i^*\big)\big]_0^i + \frac{1}{i_{max}} * [(i_{max} + m_I * i) * i^*]_i^{i_{max}} \\
&= \frac{1}{i_{max}} * [i_{max} * i + m_I * (i^2 - i^2)] + \frac{1}{i_{max}} * (i_{max} - i) * (i_{max} + m_I * i) \\
&= \frac{i}{i_{max}} * i_{max} + \left(1 - \frac{i}{i_{max}}\right) * (i_{max} + m_I * i) \\
&= i_{max} + m_I * \left(i - \frac{i^2}{i_{max}}\right)
\end{aligned} \tag{36}$$

with $m_I > 0$.

Given this, the investor seeks to maximize their expected payoff $\pi^e_{I;hoec}(i)$ within the investment's limits ($i \in [0, i_{max}]$). The resulting investment is this type of *investor's best response* $i_{BR;hoec}$. Thus, maximizing the given function of the investor's expected payoff $\pi^e_{I;hoec}(i)$ with respect to the investment i, leads to:

$$\begin{aligned} \max_i\left(\pi^e_{I;hoec}(i)\right) \quad &\rightarrow \quad \frac{\partial \pi^e_{I;hoec}(i)}{\partial i} \stackrel{!}{=} 0 \\ &\leftrightarrow \quad \frac{\partial\left(i_{max} + m_I * \left(i - \frac{i^2}{i_{max}}\right)\right)}{\partial i} = 0 \\ &\leftrightarrow \quad m_I - \frac{2 * m_I * i}{i_{max}} = 0 \\ &\leftrightarrow \quad i_{BR;hoec} = \frac{i_{max}}{2} \end{aligned} \tag{37}$$

with $m_I > 0$.

This means that this type of investor maximizes their expected payoff $\pi^e_{I;hoec}(i)$ by investing half of the maximal investment i_{max}. For this best response $i_{BR;hoec}$ the following payoff $\pi^e_{I;BR;hoec}$ is expected:

$$\begin{aligned} \pi^e_{I;BR;hoec} &= \pi^e_{I;hoec}\left(i = i_{BR;hoec}\right) \\ &= i_{max} + m_I * \left(i_{BR;hoec} - \frac{{i_{BR;hoec}}^2}{i_{max}}\right) \\ &= i_{max} + m_I * \left(\frac{i_{max}}{2} - \left(\frac{i_{max}}{2}\right)^2 * \frac{1}{i_{max}}\right) \\ &= i_{max} * \left(1 + \frac{m_I}{4}\right) \end{aligned} \tag{38}$$

with $m_I > 0$.

In short, the *homo oeconomicus* type of investor maximizes their payoff by making an investment i that equals half of the maximal investment i_{max}, regardless of the previously received advice a ($\forall a$: $i = i_{BR;hoec} = \frac{i_{max}}{2}$). Hence, this type of investor's best response $i_{BR;hoec}$ is unique. This makes the investor's best response $i_{BR;hoec}$ a strictly dominant strategy.

Anticipating this, the *homo oeconomicus* type of advisor knows that their advice a will not impact the investment i. Since the advisor's payoff $\pi_A(i)$ depends solely on the investment i, this leaves them with no means to influence their payoff $\pi_A(i)$. Thus, an advisor that only values their monetary payoff $\pi_A(i)$ will give any random advice a between 0 and the maximal investment i_{max} with the same probability ($a \in [0, i_{max}]$). Therefore, this type of *advisor's set of best responses* $S_{A;BR;hoec}$ consists of all possible advice numbers. This implies:

$$S_{A;BR;hoec} = \{a | a \in [0, i_{max}]\}. \tag{39}$$

As the investor will invest equal to their unique best response $i_{BR;hoec}$, this type of advisor will expect a payoff $\pi^{e}_{A;BR;hoec}$ that amounts to:

$$\pi^{e}_{A;BR;hoec} = \pi_A\big(i = i_{BR;hoec}\big) = m_A * i_{BR;hoec} = i_{max} * \frac{m_A}{2} \tag{40}$$

with $m_A > 0$.

Since all advice a will result in this same expected payoff $\pi^{e}_{A;BR;hoec}$, every advice $a \in [0, i_{max}]$ is a weakly dominant strategy for the *homo oeconomicus* type of advisor.

Finally, when both players solely care for their monetary payoffs, the *set of game theoretical equilibria* $S_{equ;hoec}$ for the CDG can be defined as:

$$S_{equ;hoec} = \left\{(a^{s}_{hoec}, i^{s}_{hoec}) \,\middle|\, a^{s}_{hoec} \in [0, i_{max}] \wedge i^{s}_{hoec} = \tfrac{i_{max}}{2}\right\}. \tag{41}$$

B.1.2. Monetary and non-monetary motivation

In this section, I derive the set of game theoretical equilibria of the CDG with rational players who care for more than their monetary payoffs. Since the CDG is based fundamentally on honesty and trust, it is reasonable to assume that the players in this game would assign a value to these two traits. It bears repeating that it has been theorized that our internal value system rewards honest behavior positively and dishonest behavior negatively for various reasons (e.g., Battigalli et al., 2013; Charness & Dufwenberg, 2006; Mazar et al., 2008; Vanberg, 2008). Since there are many studies in support of this idea, I assume that players have a preference for honesty. Moreover, on an interpersonal level, trust is related to positive feelings (Barefoot et al., 1998; Kuroki, 2011), while mistrust can lead to negative ones (Gurtman, 1992). For that reason, I assume that players have a preference for trust. In order to specify these *homines morales* (Alger & Weibull, 2013), I introduce different types for both players: firstly, the advisor's type that depends on their preference for honesty and, secondly, the investor's type that depends on their preference for trust.

For the advisor, this is modeled in such a way that he or she incurs *moral costs of lying* $C_L(a)$ from giving untruthful advice ($L \neq 0$). These costs $C_L(a)$ increase monotonously with the absolute value of the percentage extent of lying ($|L|$) and, thus, are a function of their advice a in relation to the predefined optimal investment i^*.[116] However, the nature of this function is specified by the advisor's type.

To be more exact, the advisor's moral costs of lying $C_L(a)$…

…become zero if the advisor behaves completely truthfully by giving advice a equal to the optimal investment i^*: $a = i^* \leftrightarrow L = 0 \rightarrow C_L(a = i^*) = 0$.

…are never negative: $C_L(a) \geq 0$.

…are a continuous function and increase monotonously with the absolute value of the percentage extent of lying $\left(|L| = \left|\frac{a-i^*}{i^*}\right|\right)$: $\frac{\partial(C_L(a))}{\partial a} = \begin{cases} a < i^*: & \leq 0 \\ a > i^*: & \geq 0 \end{cases}$.

With that, the advisor's utility $U_A(a, i)$ can be defined as the following function of their given advice a and the investment i:

$$U_A(a, i) = \pi_A(i) - C_L(a). \tag{42}$$

Analogous to the advisor, the investor's preference for trust is modeled in such a way that he or she suffers from engaging in mistrusting behavior ($\bar{T} \neq 0$). These *costs of mistrust* $C_{\bar{T}}(a, i)$ are assumed to increase monotonously with the absolute value of the investor's percentage extent of mistrust ($|\bar{T}|$). Hence, they are a function of the received advice a and the investment i. Again, the nature of this function is specified by the investor's type.

[116] This is in line with Lundquist et al. (2009) who find that the aversion to lying increases with the size of the lie. Moreover, I argue that the percentage extent of lying (L) in the CDG measures a combination of Gneezy et al.'s (2018) three dimensions of the size of the lie that determine the intrinsic costs of lying. To begin with, the *outcome dimension* (i.e., the difference between the given advice and the true optimal investment) increases continuously with the given advice by design.
Suppose now that the advisor believes that the investor will use their advice as a reference point for the investment (I will argue in favor of this assumption in more detail later). Under this assumption, and since the advisor's payoff is designed as a linear function of the investment, the advisor's expectation towards their own payoff should be strongly connected to their given piece of advice. Thus, the *payoff dimension* (i.e., the advisor's expected monetary gains from lying) can also be expected to increase with the given advice.
Finally, the advisor knows that their lying behavior can be observed ex post by the experimenters. Therefore, according to Gneezy et al. (2018), lying should always lead to the lowest possible social identity. However, I argue that in my particular experimental design the advisor will care more about how he or she is perceived by the other player (i.e., the investor) than by the experimenters. Since each value of the true optimal investment can come up with the same probability, the investor has no way to know for sure whether a received piece of advice is a lie. However, as the advisor has a monetary incentive to advise an excessive investment, it is reasonable to assume that the higher the advice, the higher is the likelihood that the investor perceives it as dishonest. For this reason, with a given optimal investment, the advisor's concerns about how he or she is perceived by the investor should increase with the extent to which he or she lies by overstating the value of the true optimum. This implies that the *likelihood dimension* of the size of the lie, which reflects concerns about one's social identity (i.e., the advisor's concerns about how he or she is perceived by others), should also be connected with the advisor's extent of lying.

More precisely, the investor's costs of mistrust $C_{\bar{T}}(a,i)$…

…become zero if the investor behaves completely trusting by making an investment i equal to the advice a: $i = a \leftrightarrow \bar{T} = 0 \rightarrow C_{\bar{T}}(a, i = a) = 0$.

…are never negative: $C_{\bar{T}}(a,i) \geq 0$.

…are a continuous function and increase monotonously with the absolute value of the percentage extent of mistrust $\left(|\bar{T}| = \left|\frac{i-a}{a}\right|\right)$: $\frac{\partial(C_{\bar{T}}(a,i))}{\partial i} = \begin{cases} i < a: & \leq 0 \\ i > a: & \geq 0 \end{cases}$.

Based on this, the investor's utility $U_I(a,i)$ can be defined as the following function of the advice a and their investment i:

$$U_I(a,i) = \pi_I(i) - C_{\bar{T}}(a,i). \tag{43}$$

Also note that each player's type is only known to themselves. However, it is assumed that the prior probability distributions over all possible realizations of both players' types, i.e., over their possible cost functions $C_L(a)$ and $C_{\bar{T}}(a,i)$, are common knowledge.[117]

In summary, up to this point, the following is given:

- i_{max}: maximal investment,
- i^*: optimal investment: This investment is randomly determined by a uniform distribution ($P([0, i^*]) = \frac{i^*}{i_{max}}$ with $i^* \in [0, i_{max}]$).
- a: advice with: $a \in [0, i_{max}]$,
- i: investment with: $i \in [0, i_{max}]$,
- i_{guess}: advisor's guess about the investment i with: $i_{guess} \in [0, i_{max}]$,
- i^*_{guess}: investor's guess about the optimal investment i^* with: $i^*_{guess} \in [0, i_{max}]$,
- $\pi_A(i)$: advisor's monetary payoff function with:
  $\pi_A(i) = m_A * i$ with $m_A > 0$,
- $\pi_I(i)$: investor's monetary payoff function with:
  $\pi_I(i) = \begin{cases} i \leq i^*: & i_{max} + m_I * i \\ i > i^*: & i_{max} + m_I * (2 * i^* - i) \end{cases}$ with $m_I > 0$,
- L: percentage extent of lying with: $L = \frac{a - i^*}{i^*}$ with $i^* > 0$,

[117] Any further assumptions about the prior probability distributions of both players' types would be rather arbitrary and, thus, I do not believe that specifying these distributions would serve the purpose of the paper in Chapter 3. However, in the results section of that paper, I empirically analyze the distribution of pursued strategies. With that in mind, at this point, I only assume that the prior probability distributions of player preferences for honesty and trust are common knowledge among the players, since this enables them to pursue their equilibrium strategies.

- L_{guess}: percentage extent of suspected lying with: $L_{guess} = \frac{a-i^*_{guess}}{i^*_{guess}}$ with $i^*_{guess} > 0$,
- $\bar{T}$: percentage extent of mistrust with: $\bar{T} = \frac{i-a}{a}$ with $a > 0$,
- $\bar{T}_{guess}$: percentage extent of expected mistrust with: $\bar{T}_{guess} = \frac{i_{guess}-a}{a}$ with $a > 0$,
- $C_L(a)$: advisor's moral costs of lying: These costs are a function of the advice a in relation to the predefined optimal investment i^*. They represent the advisor's preference for honesty. The exact nature of this function is determined by the advisor's type.
- $C_{\bar{T}}(a,i)$: investor's costs of mistrust: These costs are a function of the investment i in relation to the received advice a. They represent the investor's preference for trust. The exact nature of this function is determined by the investor's type.
- $U_A(a,i)$: advisor's utility function with: $U_A(a,i) = \pi_A(i) - C_L(a)$, and
- $U_I(a,i)$: investor's utility function with: $U_I(a,i) = \pi_I(i) - C_{\bar{T}}(a,i)$.

Moreover, I assume that the advisor and the investor aim to maximize their utility ($U_A(a,i)$ and $U_I(a,i)$, respectively). Based on this, I will first specify the advisor's set of best responses $S_{A;BR}$ (*B.1.2.1.*). Secondly, I will derive the investor's set of best responses $S_{I;BR}$ (*B.1.2.2.*). Thirdly, I will identify the set of game theoretical equilibria S_{equ} (*B.1.2.3.*). Finally, I will outline some implications for the impact of both players' beliefs on their behavior in equilibrium (*B.1.2.4.*).

B.1.2.1. The advisor's set of best responses $S_{A;BR}$

To begin with, the advisor aims to maximize their utility

$$\begin{aligned} U_A(a,i) &= \pi_A(i) - C_L(a) \\ &= m_A * i - C_L(a) \end{aligned} \tag{44}$$

with $m_A > 0$.

On the one hand, the moral costs of lying $C_L(a)$ are known to the advisor for all possible advice a, as he or she knows the value of the optimal investment i^*. On the other hand, the advisor's monetary payoff $\pi_A(i)$ is uncertain to them, since he or she does not know which investment i the investor will make. However, the prior probability distribution over all possible realizations of investor types, i.e., over all possible cost functions $C_{\bar{T}}(a,i)$, is common knowledge. Thus, the advisor is aware of the fact that the investor might have some preference for trust. Therefore, he or she knows that the investor could follow their advice a or at least use it as a reference point. This allows the advisor to give strategic advice a. With that, there is no reason why he or she would lie by giving a piece of advice a that understates the true value of the optimal investment i^* ($a < i^* \leftrightarrow L < 0$), since this would potentially lead to moral costs of lying ($C_L(a) \geq 0$) while also potentially reducing their monetary payoff $\pi_A(i)$. As a consequence, he or she would either give truthful advice a or lie by overstating the optimal investment i^* to get the investor to overinvest ($a \geq i^* \leftrightarrow L \geq 0$).[118] This implies that the advisor's set of best responses $S_{A;BR}$ consists only of advice a between the optimal i^* and the maximal investment i_{max}. This yields:

$$\forall a \in S_{A;BR}: a \in [i^*, i_{max}]. \tag{45}$$

As both players know this, the investor can be expected to make an investment i equal to or below the received advice a ($i \leq a \leftrightarrow \bar{T} \leq 0$). This in turn is known to the advisor. As a result, he or she can form their first-order beliefs about the investor's mistrust accordingly ($i_{guess} \leq a \leftrightarrow \bar{T}_{guess} \leq 0$). Based on this, the advisor can estimate their utility $U_A(a,i)$ by consulting their beliefs about the investor's type and making a guess i_{guess} on the investment i. This leads to the following function for the *advisor's expected utility* $U_A^e(a, i_{guess})$:

$$\begin{aligned} U_A^e(a, i_{guess}) &= U_A(a, i = i_{guess}) \\ &= \pi_A(i = i_{guess}) - C_L(a) \\ &= m_A * i_{guess} - C_L(a) \end{aligned} \tag{46}$$

with $m_A > 0$.

Furthermore, it should be remembered that the guessed investment i_{guess} can be expressed by the percentage extent of expected mistrust $\bar{T}_{guess}$ as follows:

$$\bar{T}_{guess} = \frac{i_{guess} - a}{a} \quad \leftrightarrow \quad i_{guess} = (\bar{T}_{guess} + 1) * a \tag{47}$$

with $a > 0$ and $\bar{T}_{guess} \leq 0$.

Note that the guessed investment i_{guess} depends on the advice a, since the investor is expected to use the advice a as a reference point for the investment i. In particular, the guessed investment i_{guess} should increase monotonously with the given advice a and become zero if the advice a is zero. Beyond that, the percentage extent of expected mistrust $\bar{T}_{guess}$ also depends on the advice a by definition.

On this basis, the advisor's expected utility $U_A^e\left(a, \bar{T}_{guess}(a)\right)$ can be expressed as a function of their first-order beliefs about the investor's mistrusting behavior $\bar{T}_{guess}(a)$ and the advice a as follows:

[118] Note that an honest type of advisor incurs higher moral costs of lying than a dishonest type of advisor. This means that a completely honest advisor gives advice a equal to the optimal investment i^* ($a = i^* \leftrightarrow L = 0$), whereas a completely dishonest advisor tries to maximize their monetary payoff $\pi_A(i)$ by giving advice a above the optimal investment i^* if necessary ($a > i^* \leftrightarrow L > 0$).

$$U_A^e\left(a, \bar{T}_{guess}(a)\right) = m_A * i_{guess} - C_L(a)$$
$$= m_A * \left(\bar{T}_{guess}(a) + 1\right) * a - C_L(a) \quad (48)$$

with $m_A > 0$ and $\bar{T}_{guess} \leq 0$.

Here the *homo moralis* type of advisor faces a trade-off between maximizing their estimated monetary payoff $\pi_A(i)$ (with: $i = \left(\bar{T}_{guess} + 1\right) * a$) and reducing their moral costs of lying $C_L(a)$ – or, in other words, between the monetary incentive of lying and their preference for honesty. This trade-off can be solved by maximizing the expected utility $U_A^e\left(a, \bar{T}_{guess}(a)\right)$ with respect to the advice $a \in [i^*, i_{max}]$. This yields:

$$\max_a\left(U_A^e\left(a, \bar{T}_{guess}(a)\right) \middle| \ a \in [i^*, i_{max}]\right) \quad \rightarrow \quad \frac{\partial U_A^e\left(a, \bar{T}_{guess}(a)\right)}{\partial a} \overset{!}{=} 0$$
$$\leftrightarrow \quad \frac{\partial\left(m_A * \left(\bar{T}_{guess}(a) + 1\right) * a - C_L(a)\right)}{\partial a} = 0 \quad (49)$$
$$\leftrightarrow \quad \frac{\partial\left(C_L(a)\right)}{\partial a} = m_A * \frac{\partial\left(\left(\bar{T}_{guess}(a) + 1\right) * a\right)}{\partial a}$$

with $m_A > 0$ and $\bar{T}_{guess} \leq 0$.

This means that any interior solution to the *advisor's maximization problem* must meet the condition that the derivative of the advisor's moral costs of lying $C_L(a)$ must be equal to the term $m_A * \frac{\partial\left(\left(\bar{T}_{guess}(a)+1\right)*a\right)}{\partial a}$. Thus, any interior solution corresponds to giving advice a (between the optimal i^* and the maximal investment i_{max}) such that the advisor's preference for honesty (i.e., their sensitivity to change of their moral costs of lying) is balanced in a specific way with their monetary incentive (i.e., the rate at which their payoff increases with the size of the investment) and their belief about the other player's type (i.e., the sensitivity to change of their guess about the investor's mistrust in combination with the advice). Note, however, that this maximization problem could also have a boundary solution (i.e., completely honest or completely dishonest behavior). For this reason, it is also possible that either giving completely honest advice a, equal to the optimal investment i^* ($a = i^*$), or giving the highest possible advice a, equal to the maximal investment i_{max} ($a = i_{max}$), solves this problem.

It follows that all pieces of advice a that are included in the advisor's set of best responses $S_{A;BR}$ must fulfill the following condition:

$$\forall a \in S_{A;BR}: \quad \frac{\partial\left(C_L(a)\right)}{\partial a} = m_A * \frac{\partial\left(\left(\bar{T}_{guess}(a)+1\right)*a\right)}{\partial a} \lor a = i^* \lor a = i_{max}. \quad (50)$$

Based on this, the *advisor's set of best responses* $S_{A;BR}$ can be defined as:

$$S_{A;BR} = \left\{ a^s \,\middle|\, a^s, a^{s\prime} \in \left\{ a \in [i^*, i_{max}] \,\middle|\, \frac{\partial\left(C_L(a)\right)}{\partial a} = m_A * \frac{\partial\left(\left(\bar{T}_{guess}(a)+1\right)*a\right)}{\partial a} \vee a = i^* \vee a = i_{max} \right\} \wedge \forall a^{s\prime}: U_A^e(a^s) \geq U_A^e(a^{s\prime}) \right\} \tag{51}$$

with $m_A > 0$ and $\bar{T}_{guess} \leq 0$.

B.1.2.2. The investor's set of best responses $S_{I;BR}$

Analogous to the advisor, the investor aims to maximize their utility

$$\begin{aligned} U_I(a,i) &= \pi_I(i) - C_{\bar{T}}(a,i) \\ &= \begin{cases} i \leq i^*: & i_{max} + m_I * i - C_{\bar{T}}(a,i) \\ i > i^*: & i_{max} + m_I * (2 * i^* - i) - C_{\bar{T}}(a,i) \end{cases} \end{aligned} \tag{52}$$

with $m_I > 0$.

After receiving the advice a, the investor knows their costs of mistrust $C_{\bar{T}}(a,i)$ for every possible investment i. However, the investor has no way of knowing their exact monetary payoff $\pi_I(i)$ because he or she has no further information about the true value of the optimal investment i^*. Yet, when the investor receives the advice a, he or she learns the upper limit of the optimal investment i^*, since the advisor is expected to either give truthful advice a or lie by overstating the value of the optimal investment i^* ($a \geq i^* \leftrightarrow L \geq 0$).[119] As the investor anticipates this ($a \geq i_{guess}^* \leftrightarrow L_{guess} \geq 0$), he or she always makes an investment i less than or equal to their received advice a ($i \leq a \leftrightarrow \bar{T} \leq 0$).[120] In addition, there is no reason why the investor would make an investment i below their guess i_{guess}^* on the optimal investment i^*. As a consequence, the investor's set of best responses $S_{I;BR}$ consists only of investments i between their guessed optimal investment i_{guess}^* and the advice a. Hence:

$$\forall i \in S_{I;BR}: i \in \left[i_{guess}^*, a\right]. \tag{53}$$

In order to make a sound investment, the investor needs to consider the commonly known prior probability distribution over all possible realizations of advisor types, i.e., over all possible cost functions $C_L(a)$. On this basis, he or she can estimate their utility $U_I(a,i)$ by consulting their beliefs

---

[119] For that reason, receiving a low value piece of advice a means bad news for the investor. I owe that point to Johann Graf Lambsdorff.

[120] Aware of the possibility that the advisor wants to avoid lying, the investor can use the advice a as a reference point for their investment i. This means that the more trusting the investor, the more he or she follows the advice a. Thus, a completely trusting investor would exactly follow the advice a ($i = a \leftrightarrow \bar{T} = 0$), while a completely mistrusting investor would try to maximize their monetary payoff $\pi_I(i)$, most likely resulting in an investment i below the advice a ($i < a \leftrightarrow \bar{T} < 0$).

about the advisor's type and making a guess i_{guess}^* on the optimal investment i^*. This allows me to formulate the *investor's expected utility* $U_I^e(a, i, i_{guess}^*)$ as the following function of the advice a, their investment i, and their estimate of the optimal investment i_{guess}^*:

$$\begin{aligned} U_I^e(a, i, i_{guess}^*) &= \pi_I(i)_{i^*=i_{guess}^*} - C_{\bar{T}}(a, i) \\ &= \begin{cases} i \leq i_{guess}^*: & i_{max} + m_I * i - C_{\bar{T}}(a, i) \\ i > i_{guess}^*: & i_{max} + m_I * \left(2 * i_{guess}^* - i\right) - C_{\bar{T}}(a, i) \end{cases} \end{aligned} \tag{54}$$

with $m_I > 0$.

It should be remembered that the investor's guess i_{guess}^* on the location of the true optimal investment i^* can be expressed by the percentage extent of suspected lying L_{guess} as follows:

$$L_{guess} = \frac{a - i_{guess}^*}{i_{guess}^*} \quad \leftrightarrow \quad i_{guess}^* = \frac{a}{L_{guess} + 1} \tag{55}$$

with $i_{guess}^* > 0$ and $L_{guess} \geq 0$.

Based on this, the investor's expected utility $U_I^e(a, i, L_{guess})$ can be expressed as the following function of the advice a, the investment i, and the investor's first-order beliefs about the advisor's lying behavior L_{guess}:

$$\begin{aligned} U_I^e(a, i, L_{guess}) &= U_I^e\left(a, i, i_{guess}^* = \frac{a}{L_{guess} + 1}\right) \\ &= \begin{cases} i \leq \dfrac{a}{L_{guess} + 1}: & i_{max} + m_I * i - C_{\bar{T}}(a, i) \\ i > \dfrac{a}{L_{guess} + 1}: & i_{max} + m_I * \left(2 * \dfrac{a}{L_{guess} + 1} - i\right) - C_{\bar{T}}(a, i) \end{cases} \\ &= \begin{cases} i < \dfrac{a}{L_{guess} + 1}: & i_{max} + m_I * i - C_{\bar{T}}(a, i) \\ i \geq \dfrac{a}{L_{guess} + 1}: & i_{max} + m_I * \left(2 * \dfrac{a}{L_{guess} + 1} - i\right) - C_{\bar{T}}(a, i) \end{cases} \end{aligned} \tag{56}$$

with $m_I > 0$ and $L_{guess} \geq 0$.[121]

This function reflects the fact that the *homo moralis* type of investor faces a trade-off between maximizing their estimated monetary payoff $\pi_I(i)$ (with: $i^* = i_{guess}^* = \frac{a}{L_{guess}+1}$) and reducing their

[121] The last transformation of the expected utility $U_I^e(a, i, L_{guess})$ is valid because $U_I^e(a, i, L_{guess})$ is a continuous function. For $i = i_{guess}^* = \frac{a}{L_{guess}+1}$ this implies:

$$i_{max} + m_I * i - C_{\bar{T}}(a, i) = i_{max} + m_I * \left(2 * \frac{a}{L_{guess} + 1} - i\right) - C_{\bar{T}}(a, i)$$

with $m_I > 0$ and $L_{guess} \geq 0$.

costs of mistrust $C_{\bar{T}}(a,i)$ – or to put it differently – between the monetary incentive to invest optimally and their preference for trust. In order to solve this trade-off problem, the investor can maximize their expected utility $U_I^e(a,i,L_{guess})$ with regard to their investment i. Therefore, it must be considered that this function's domain is split into two regions ($i < i^*_{guess} = \frac{a}{L_{guess}+1}$ and $i \geq i^*_{guess} = \frac{a}{L_{guess}+1}$). However, compared to all possible investments within the first region ($i < i^*_{guess} = \frac{a}{L_{guess}+1}$), the investor can always increase their expected utility $U_I^e(a,i,L_{guess})$ by making an investment i equal to their guessed optimal investment i^*_{guess}, since this would not only reduce the investor's costs of mistrust $C_{\bar{T}}(a,i)$ but also increase their expected payoff (which then would be equal to: $i_{max} + m_I * i^*_{guess}$). It follows that the investor can only maximize their expected utility $U_I^e(a,i,L_{guess})$ within the second region ($i \geq i^*_{guess} = \frac{a}{L_{guess}+1}$).

Now, maximizing the investor's expected utility $U_I^e(a,i,L_{guess})$ with regard to the investment $i \in \left[\frac{a}{L_{guess}+1}, a\right]$ leads to:

$$\max_i \left(U_I^e(a,i,L_{guess}) \,\middle|\, i \in \left[\frac{a}{L_{guess}+1}, a\right] \right) \quad \rightarrow \quad \frac{\partial U_I^e(a,i,L_{guess})}{\partial i} \stackrel{!}{=} 0$$
$$\leftrightarrow \quad -m_I - \frac{\partial\left(C_{\bar{T}}(a,i)\right)}{\partial i} = 0 \qquad (57)$$
$$\leftrightarrow \quad \frac{\partial\left(C_{\bar{T}}(a,i)\right)}{\partial i} = -m_I$$

with $m_I > 0$ and $L_{guess} \geq 0$.

This means that any interior solution to the *investor's maximization problem* must meet the condition that the derivative of the investor's costs of mistrust $C_{\bar{T}}(a,i)$ must be equal to their payoff factor $-m_I$ (by which the investor's payoff $\pi_I(a,i)$ decreases when the investment i deviates from the optimum i^*). Here the former depends on the investor's type, while the latter is given by their payoff structure. As a consequence, any interior solution corresponds to making an investment i (between their guess i^*_{guess} on the optimal investment i^* and the received advice a) such that the investor's preference for trust (i.e., their sensitivity to change of their costs of mistrust) is balanced in a specific way with their monetary incentive (i.e., with the rate at which their payoff decreases when their investment deviates from its optimum). However, this maximization problem could also have a boundary solution (i.e., completely trusting or mistrusting behavior). Therefore, it could also be solved by either a completely trusting investment i, equal to the received advice a ($i = a$), or a mistrusting investment i, equal to the guessed optimal investment i^*_{guess} ($i = i^*_{guess}$).

Hence, all investments i that are contained in the investor's set of best responses $S_{I;BR}$ must meet the following condition:

$$\forall i \in S_{I;BR}: \quad \frac{\partial\left(C_{\bar{T}}(a,i)\right)}{\partial i} = -m_I \vee i = \frac{a}{L_{guess}+1} \vee i = a. \tag{58}$$

On this basis, the *investor's set of best responses* $S_{I;BR}$ can be defined as:

$$S_{I;BR} = \left\{ i^s \,\middle|\, i^s, i^{s'} \in \left\{ i \in \left[\frac{a}{L_{guess}+1}, a\right] \,\middle|\, \frac{\partial\left(C_{\bar{T}}(a,i)\right)}{\partial i} = -m_I \vee i = \frac{a}{L_{guess}+1} \vee i = a \right\} \wedge \forall i^{s'}: U_I^e(i^s) \geq U_I^e(i^{s'}) \right\} \tag{59}$$

with $m_I > 0$ and $L_{guess} \geq 0$.

B.1.2.3. The set of equilibria S_{equ}

The set of game theoretical equilibria S_{equ} occurs at the intersection of both players' sets of best responses ($S_{A;BR}$ and $S_{I;BR}$) and under the condition that both players' beliefs about each other are correct, which implies: $L = L_{guess}$ and $\bar{T} = \bar{T}_{guess}$. It can be concluded that, when both players consider not only their monetary payoffs but also value honesty and trust, the *set of game theoretical equilibria* S_{equ} for the CDG is:

$$S_{equ} = \left\{ (a^s, i^s) \,\middle|\, \left(a^s, a^{s'} \in \left\{ a \in [i^*, i_{max}] \,\middle|\, \frac{\partial\left(C_L(a)\right)}{\partial a} = m_A * \frac{\partial\left(i^s(a)\right)}{\partial a} \vee a = i^* \vee a = i_{max} \right\} \wedge \forall a^{s'}: U_A^e(a^s) \geq U_A^e(a^{s'}) \right) \wedge \left(i^s, i^{s'} \in \left\{ i \in [i^*, a^s] \,\middle|\, \frac{\partial\left(C_{\bar{T}}(a^s,i)\right)}{\partial i} = -m_I \vee i = i^* \vee i = a^s \right\} \wedge \forall i^{s'}: U_I^e(i^s) \geq U_I^e(i^{s'}) \right) \right\} \tag{60}$$

with $m_A > 0$ and $m_I > 0$.

Depending on both players' types and their beliefs about each other, there remain four possible combinations of classes of *rational strategies* that could be pursued by the advisor and the investor in equilibrium. These combinations are summarized in Table B-1.

	Strategy combinations			Outcome of the investment
	Advisor		Investor	
1	*Cooperative truth-telling*	-	*Cooperative trust*	*Optimal investment* ($F = 0$)
2	*Optimal profitable white lie*	-	*Optimal profitable mistrust*	*Optimal investment* ($F = 0$)
3	*Selfish lie*	-	*Suboptimal profitable mistrust*	*Overinvestment* ($F > 0$)
4	*Selfish lie*	-	*Benevolent trust*	*Overinvestment* ($F \gg 0$)

Table B-1. Possible equilibrium strategy combinations

It can be seen that combination 1 (i.e., mutually cooperative behavior) and combination 2 (i.e., fully equalizing behavior) result in an optimal investment ($i = i^* \leftrightarrow F = 0$). Hence, both combinations lead to the same financial outcome for both players. However, combination 2 is less efficient for players who have a preference for honesty or trust. Even though both players might prefer a more cooperative set of strategies, neither of them would benefit from a unilateral deviation from their equilibrium strategy. By contrast, combination 3 (i.e., partially advantageous behavior) and combination 4 (i.e., fully advantageous behavior) result in an overinvestment ($i > i^* \leftrightarrow F > 0$), where the advisor monetarily benefits from the investor's preference for trust. In both of these combinations the investor values trust so highly that he or she is willing to accept a financial loss in order to behave trustingly. In combination 4, the investor's preference for trust is so strong that he or she values trust entirely over additional financial gain.

B.1.2.4. Further implications

After analyzing what rational players would do in the CDG, I wish to outline some implications that arise from the set of game theoretical equilibria. Therefore, I will focus on the rational *homo moralis* types of players and discuss how their first-order beliefs about the other player would influence their behavior.

So far, the cost functions ($C_L(a)$ and $C_{\bar{T}}(a, i)$) were defined very generally. Going into more detail, I argue in favor of both *diminishing marginal costs of lying and mistrust*. It is reasonable to suppose that the advisor would suffer more from a marginal higher extent of lying if he or she originally planned to give truthful advice a than if he or she already planned to engage in a high extent of lying anyway. This is in line with Ariely (2012) and Engelmann and Fehr (2016) who argue that one finds it easier to behave dishonestly when one has already justified being dishonest to some extent. The same can be assumed in regard to the investor's preference for trust: The investor would suffer more from behaving marginally more mistrustingly if he or she originally chose to trust their advisor than if he or she already chose to mistrust the advisor.

To meet these conditions, diminishing marginal costs of lying are assumed for the advisor and diminishing marginal costs of mistrust for the investor. With that, both players' cost functions are concave with a zero point for completely truthful ($L = 0 \rightarrow C_L(a) = 0$) or, respectively, completely trusting ($\bar{T} = 0 \rightarrow C_{\bar{T}}(a, i) = 0$) behavior. This leads to the following conditions:

$$\frac{\partial}{\partial a}\left(\frac{\partial\big(C_L(a)\big)}{\partial a}\right) = \begin{cases} a < i^*: & \leq 0 \\ a > i^*: & \leq 0 \end{cases} \tag{61}$$

for the advisor's cost function $C_L(a)$ and

$$\frac{\partial}{\partial i}\left(\frac{\partial\big(C_{\bar{T}}(a, i)\big)}{\partial i}\right) = \begin{cases} i < a: & \leq 0 \\ i > a: & \leq 0 \end{cases} \tag{62}$$

for the investor's cost function $C_{\bar{T}}(a, i)$.

Based on this, the nature of both players in the previously defined set of equilibria S_{equ} can be used to draw conclusions on the impact that each player's first-order beliefs about the other player ($\bar{T}_{guess}$ or L_{guess}) have on their behavior (L or, respectively, $\bar{T}$) in equilibrium.

As shown before, for any interior solution to the *advisor*'s maximization problem in equilibrium, the following applies:

$$\frac{\partial\big(C_L(a)\big)}{\partial a} = m_A * \frac{\partial\big(i(a)\big)}{\partial a} \tag{63}$$

with $m_A > 0$.

On this basis, it can be shown that the more mistrusting the advisor believes the investor to be, the more dishonest he or she behaves. To explain this, it shall be reminded that, in the CDG, a more mistrusting type of investor responds with a higher absolute value of the percentage extent of (risk-reducing) mistrust ($|\bar{T}(a)| \uparrow$ with $\bar{T}(a) \leq 0$) to any advice ($\forall a$) that he or she receives. Anticipating this, the advisor would expect the absolute value of the investor's extent of (risk-reducing) mistrust ($\left|\bar{T}_{guess}(a)\right| \uparrow$ with $\bar{T}_{guess}(a) \leq 0$) to be higher for any advice ($\forall a$). As a result, the advisor would lie by overstating to a larger extent ($L \uparrow$ with $L \geq 0$), since:

$\lvert\bar{T}_{guess}(a)\rvert \uparrow$ (for all advice a)	$\rightarrow \quad \bar{T}_{guess}(a) \downarrow$	(since: $\bar{T}_{guess}(a) \leq 0$)
	$\rightarrow \quad i_{guess}(a) \downarrow$	(since: $i_{guess}(a) = \left(\bar{T}_{guess}(a) + 1\right) * a$)
	$\rightarrow \quad \frac{\partial\left(i_{guess}(a)\right)}{\partial a} \downarrow$	(since: $i_{guess}(a)$ becomes zero for $a = 0$, is concave, and increases monotonously.[122] Thus, if $\forall a$: $i_{guess}(a) \downarrow$, this function must be flatter. It follows that, for any increment of the advice a, the increment of $i_{guess}(a)$ must be lower.)
	$\rightarrow \quad \frac{\partial\left(i(a)\right)}{\partial a} \downarrow$	(since: $i_{guess} = i$ due to correct beliefs in equilibrium)
	$\rightarrow \quad \left(m_A * \frac{\partial\left(i(a)\right)}{\partial a}\right) \downarrow$	(since: $m_A > 0$)
	$\rightarrow \quad \frac{\partial\left(C_L(a)\right)}{\partial a} \downarrow$	(since: $\frac{\partial\left(C_L(a)\right)}{\partial a} = m_A * \frac{\partial\left(i(a)\right)}{\partial a}$ must be fulfilled for any interior solution to the advisor's maximization problem in equilibrium)
	$\rightarrow \quad C_L(a)_{\text{accepted by advisor}} \uparrow$	(since: $C_L(a)$ is concave and increases monotonously for $a \geq i^*$)
	$\rightarrow \quad a \uparrow$	(since: $C_L(a)$ increases monotonously for $a \geq i^*$)
	$\rightarrow \quad (a - i^*) \uparrow$	(since: i^* is constant)
	$\rightarrow \quad L \uparrow$	(since: $L \geq 0 \leftrightarrow a \geq i^*$).

This means that, in equilibrium, higher expectations of being mistrusted (i.e., a larger extent of expected mistrust $\lvert\bar{T}_{guess}\rvert$) make the advisor engage in more dishonest behavior (i.e., a larger extent of lying L).

Turning to the *investor*, for any interior solution to their maximization problem in equilibrium, the following condition must be fulfilled:

$$\frac{\partial\left(C_{\bar{T}}(a,i)\right)}{\partial i} = -m_I \tag{64}$$

with $m_I > 0$.

This equation has an intuitive interpretation: The higher the investor's monetary incentive to invest optimally ($m_I \uparrow$), the higher costs of mistrust $C_{\bar{T}}(a,i)$ he or she is willing to accept. This in turn would result in a lower (and therefore more risk-reducing) investment i, since:

[122] It should be reminded that the advisor's guess i_{guess} on the investment i depends on the advice a as follows: Since the investor can be expected to use the advice a as a reference point for the investment i, the guessed investment i_{guess} increases monotonously with the given advice a and becomes zero if the advice a is zero. In addition, it is known to the investor that the advisor has an incentive to lie by overstating ($L > 0$). For this reason, the higher the advice a, the less the advisor should expect the investor to be influenced by an increment of the advice a. Thus, the advisor's guess i_{guess} on the investment i can be assumed to be a concave function of the advice a.

$m_I \uparrow \quad \rightarrow \quad (-m_I) \downarrow$

$\rightarrow \quad \frac{\partial\left(C_{\bar{T}}(a,i)\right)}{\partial i} \downarrow$ (since: $\frac{\partial\left(C_{\bar{T}}(a,i)\right)}{\partial i} = -m_I$ must be fulfilled for any interior solution to the investor's maximization problem in equilibrium)

$\rightarrow \quad C_{\bar{T}}(a,i)_{\text{accepted by investor}} \uparrow$ (since: $C_{\bar{T}}(a,i)$ is concave and decreases monotonously for $i \leq a$)

$\rightarrow \quad i \downarrow$ (since: $C_{\bar{T}}(a,i)$ decreases monotonously for $i \leq a$).

As shown before, the lower limit for the investment i in equilibrium corresponds to the investor's guess i^*_{guess} on the optimal investment i^*, which is equal to:

$$i^*_{guess} = \frac{a}{L_{guess} + 1} \tag{65}$$

with $L_{guess} \geq 0$.

This lower limit depends on the received advice a and the investor's belief about the advisor's dishonesty (i.e., the extent L_{guess} to which the investor suspects their advisor to have lied). It follows that the more the investor suspects their advisor to lie by overstating ($L_{guess} \uparrow$ with $L_{guess} \geq 0$), the lower investments i he or she potentially considered in equilibrium, since:

$L_{guess} \uparrow \quad \rightarrow \quad \left(\frac{a}{L_{guess} + 1}\right) \downarrow$ (since: $a \geq 0$ and $L_{guess} \geq 0$)

$\rightarrow \quad i^*_{guess} \downarrow$ (since: $i^*_{guess} = \frac{a}{L_{guess}+1}$)

$\rightarrow \quad \min(i) \downarrow$ (since: $i \in \left[i^*_{guess}, a\right]$)

$\rightarrow \quad \max(|\bar{T}|) \uparrow$ (since: $i \leq a \leftrightarrow \bar{T} \leq 0$ and $\bar{T} = \frac{i-a}{a}$).

In other words, in equilibrium a stronger suspicion of being lied to (expressed in L_{guess}) makes the investor consider more mistrusting strategies (i.e., a higher possible extent of risk-reducing mistrust $|\bar{T}|$ with $\bar{T} \leq 0$). However, their decision on the investment i ultimately depends on the relation between their preference for trust and their monetary incentive to invest optimally.

B.2. Materials

The following subsections present a translation of the pre-game instructions (*B.2.1.*) and input screens (*B.2.2.*) for the CDG in *classEx*.[123] Original German materials are available upon request.

B.2.1. Pre-game instructions

B.2.1.1. Instructions for the advisor (i.e., the sender)

class EX

Your alias: Advisor-1-203970

Please read these instructions carefully before the experiment starts.

If you have any questions or concerns, please raise your hand. The experimenters will answer your questions individually and in private.

Experimental procedure:

- This experiment consists of **10 rounds**. All rounds are independent from each other.
- After finishing all rounds please fill out the anonymous **questionnaire**.
- Afterwards, you will receive your **payoff**, which depends on your decisions in the experiment. Therefore, you can earn coins in each round. At the end of the experiment, one round will be selected randomly. This round will be used to determine your payoff (all others rounds will not be considered).
- The **number of coins** you earned in the selected round will be exchanged to Euros by the following rate: 100 coins = 8€ (the payoffs will be rounded up to the nearest 10 cent value).

Your role:

- In all rounds you take the role of an **advisor**.
- In each round you are randomly assigned to another player who takes the role of an investor. Note that you will **never be matched with the same player twice**.
- Note that you are **completely anonymous** throughout the entire experiment and will not be informed about the identity of the other players at any time.

Figure B-1. Instructions for the advisor – Part 1/5

[123] Technical instructions regarding the use of *classEx* on mobile phones were presented separately from these instructions and are omitted here. For more details on *classEx*, see Giamattei and Lambsdorff (2019).

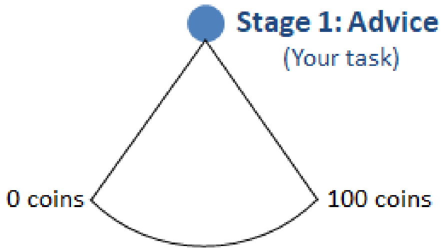
<u>Procedure within a round</u>:

Each round consists of two stages:

(1) Your task:

Report the value of the optimal investment, which maximizes the investor's payoff in this round, to the investor.

(value between 0 and 100 coins)



(2) Task of your investor:

The investor decides on the amount of the investment.

(value between 0 and 100 coins)

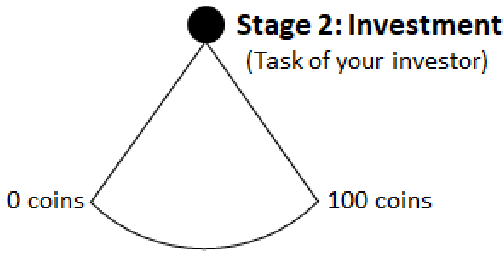


Figure B-2. Instructions for the advisor – Part 2/5

Amount of the investment:

- In each round the investor starts with 100 coins. After receiving your advice, the **investor decides how many coins between 0 and 100** he or she wants to invest.
- The investor does not know which investment maximizes their payoff.
- The investors will be informed about the outcomes of their investments of all 10 rounds only at the end of the experiment.

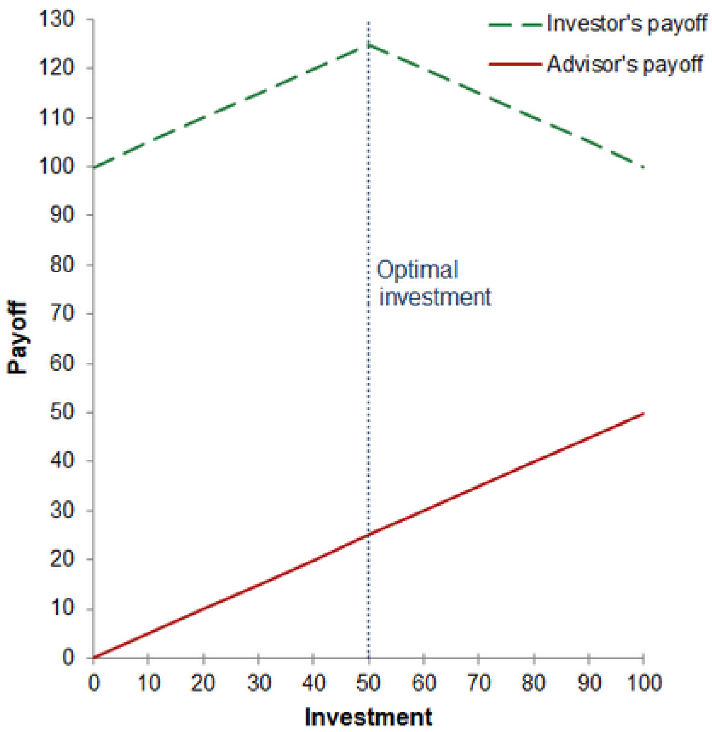
Optimal investment:

- **In each round** there is **one** optimal investment that maximizes the investor's payoff.
- The value of the optimal investment is determined randomly. Therefore, it is likely to **change between rounds**. This is known to the investor.
- You will be informed about the value of the optimal investment at the beginning of each round.
- The **investor will get no further information** about the value of the optimal investment **except your advice**.

Figure B-3. Instructions for the advisor – Part 3/5

Investment conditions:

- **Your win from the investment:** The higher the amount of the investment, the higher is your payoff.
- **The investor can win or lose from the investment:** The closer the amount of the investment to the value of the optimal investment, the higher is the investor's payoff.



[Note to the reader, not to the subjects: In order to visualize that the value of the optimal investment could be located anywhere between 0 and 100 coins, I used an animated version of this figure in which the value of the optimal investment and the corresponding peak of the investor's payoff function were moving across the screen from left to right and back again.]

The true value of the optimal investment is located somewhere between 0 and 100 coins in each round.

Figure B-4. Instructions for the advisor – Part 4/5

Advice:

Your task in each round: Report the value of the optimal investment, which maximizes the investor's payoff in this round, to the investor.

- Note that you can either inform your investor *honestly* about the true value of the optimal investment or *lie* by giving false advice.
- You will be informed about your investors' amounts of investments of all 10 rounds only at the end of the experiment.

Bear in mind that your **investor** will **never be the same person** in two different rounds.

Figure B-5. Instructions for the advisor – Part 5/5

B.2.1.2. Instructions for the investor (i.e., the receiver)

class EX

Your alias: Investor-1-203969

Please read these instructions carefully before the experiment starts.

If you have any questions or concerns, please raise your hand. The experimenters will answer your questions individually and in private.

Experimental procedure:

- This experiment consists of **10 rounds**. All rounds are independent from each other.
- After finishing all rounds please fill out the anonymous **questionnaire**.
- Afterwards, you will receive your **payoff**, which depends on your decisions in the experiment. Therefore, you can earn coins in each round. At the end of the experiment, one round will be selected randomly. This round will be used to determine your payoff (all others rounds will not be considered).
- The **number of coins** you earned in the selected round will be exchanged to Euros by the following rate: 100 coins = 8€ (the payoffs will be rounded up to the nearest 10 cent value).

Your role:

- In all rounds you take the role of an **investor**.
- In each round you are randomly assigned to another player who takes the role of an advisor. Note that you will **never be matched with the same player twice.**
- Note that you are **completely anonymous** throughout the entire experiment and will not be informed about the identity of the other players at any time.

Figure B-6. Instructions for the investor – Part 1/5

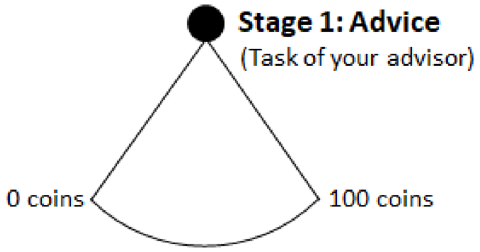
Procedure within a round:

Each round consists of two stages:

(1) Task of your advisor:

The advisor is instructed to report the value of the optimal investment, which maximizes your payoff in this round, to you.

(value between 0 and 100 coins)



(2) Your task:

Decide on the amount of your investment.

(value between 0 and 100 coins)

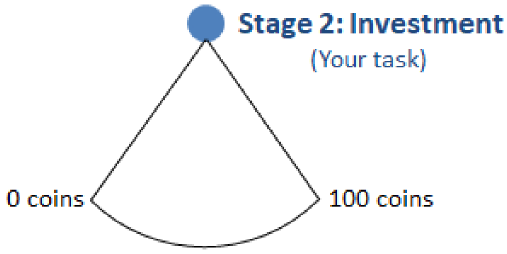


Figure B-7. Instructions for the investor – Part 2/5

Amount of the investment:

Your task in each round: After receiving advice from your advisor, decide how many coins you want to invest in this round.

- In each round you start with 100 coins. Therefore, you can invest **between 0 and 100 coins** in each round.
- Coins are not transferable between rounds.
- You will be informed about the outcomes of your investments of all 10 rounds only at the end of the experiment.

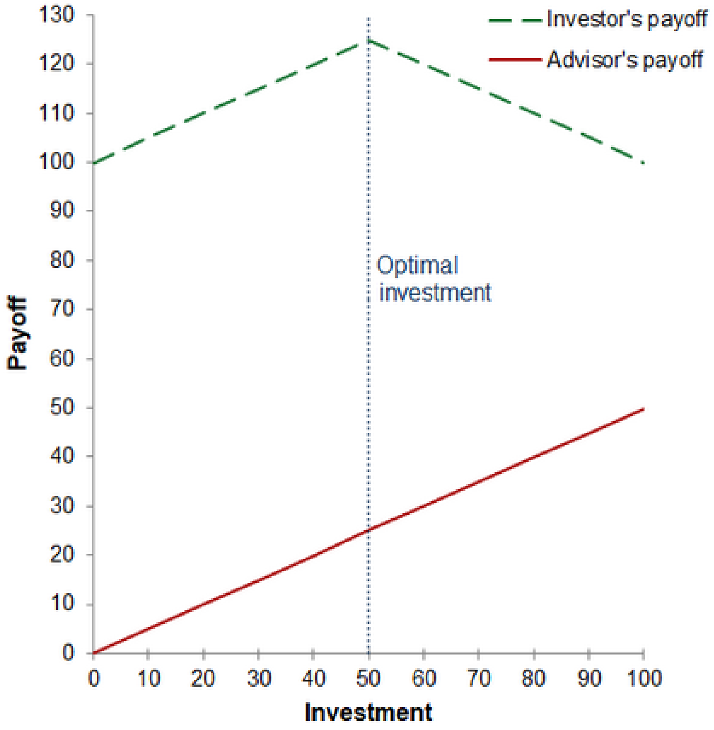
Optimal investment:

- **In each round** there is **one** optimal investment that maximizes your payoff.
- The value of the optimal investment is determined randomly. Therefore, it is likely to **change between rounds**.
- Your advisor will be informed about the value of the optimal investment at the beginning of each round.

Figure B-8. Instructions for the investor – Part 3/5

Investment conditions:

- **The advisor wins from your investment:** The higher the amount of your investment, the higher is the advisor's payoff.
- **You can win or lose from your investment:** The closer the amount of your investment to the value of the optimal investment, the higher is your payoff.



[*Note to the reader, not to the subjects: In order to visualize that the value of the optimal investment could be located anywhere between 0 and 100 coins, I used an animated version of this figure in which the value of the optimal investment and the corresponding peak of the investor's payoff function were moving across the screen from left to right and back again.*]

The true value of the optimal investment is located somewhere between 0 and 100 coins in each round.

Figure B-9. Instructions for the investor – Part 4/5

Advice:

In each round your advisor has full information about the investment conditions:

- The advisor knows the value of the optimal investment and is instructed to report this information to you.
- Note that he or she can either inform you *honestly* about the true value of the optimal investment or *lie* by giving false advice.
- You will be informed about the true values of the optimal investments of all 10 rounds only at the end of the experiment.

Bear in mind that your **advisor** will **never be the same person** in two different rounds.

Figure B-10. Instructions for the investor – Part 5/5

B.2.2. Input screens (for one round of the CDG)

B.2.2.1. Input screen for the advisor (i.e., the sender)

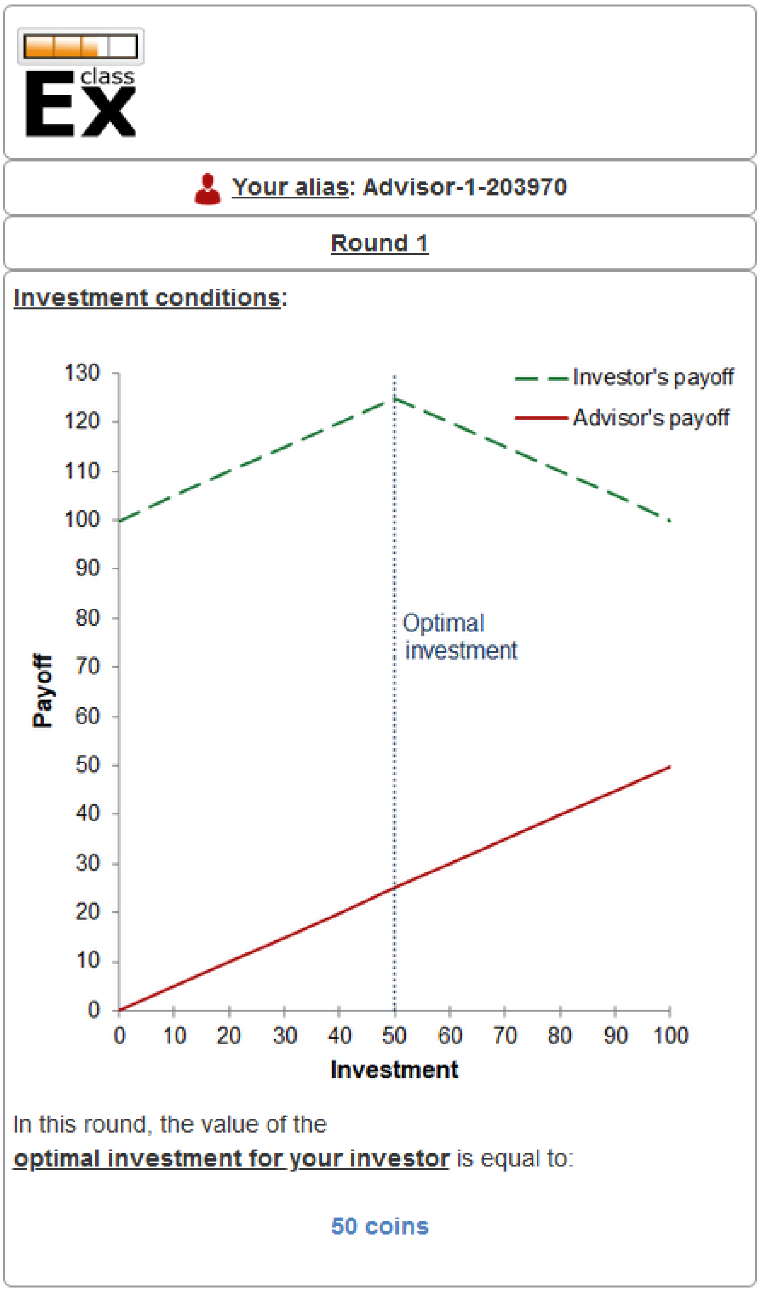


Figure B-11. Input screen for the advisor – Part 1/2

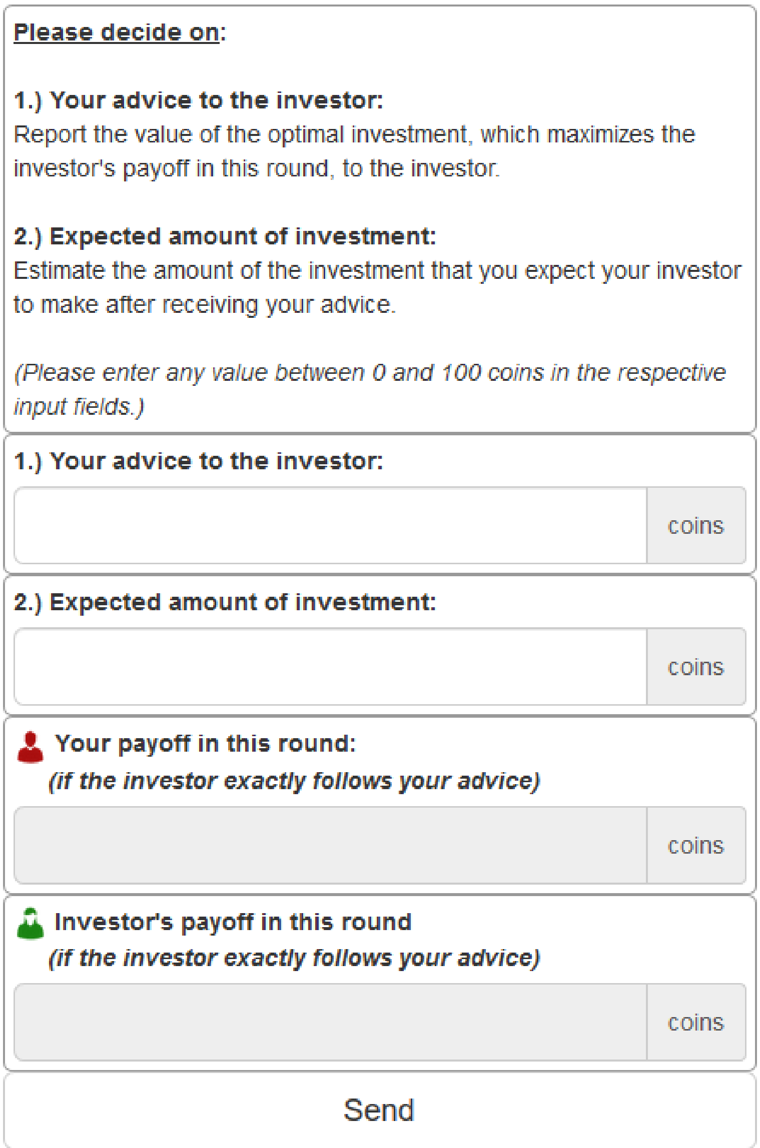


Figure B-12. Input screen for the advisor – Part 2/2

Note. The two greyed-out fields in Figure B-12 automatically indicate the payoffs that both players would receive if the investor exactly followed the advice number that the advisor has entered in the first input field.

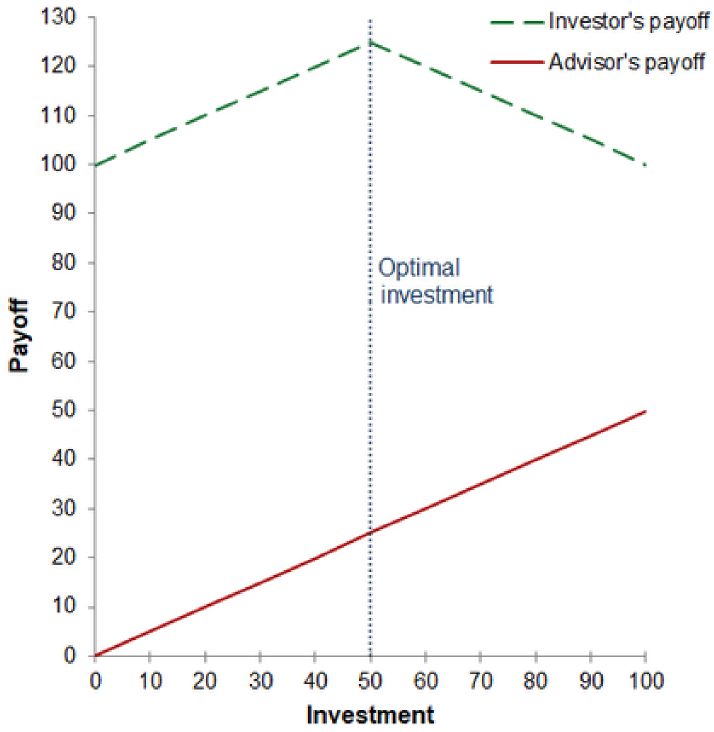
B.2.2.2. Input screen for the investor (i.e., the receiver)

Your alias: Investor-1-203969

Round 1

Investment conditions:

In each round, the true value of the optimal investment for you is located somewhere between 0 and 100 coins.



[Note to the reader, not to the subjects: In order to visualize that the value of the optimal investment could be located anywhere between 0 and 100 coins, I used an animated version of this figure in which the value of the optimal investment and the corresponding peak of the investor's payoff function were moving across the screen from left to right and back again.]

Your advisor has reported that the **value of the optimal investment**, which maximizes your payoff in this round, is equal to:

50 coins

Figure B-13. Input screen for the investor – Part 1/2

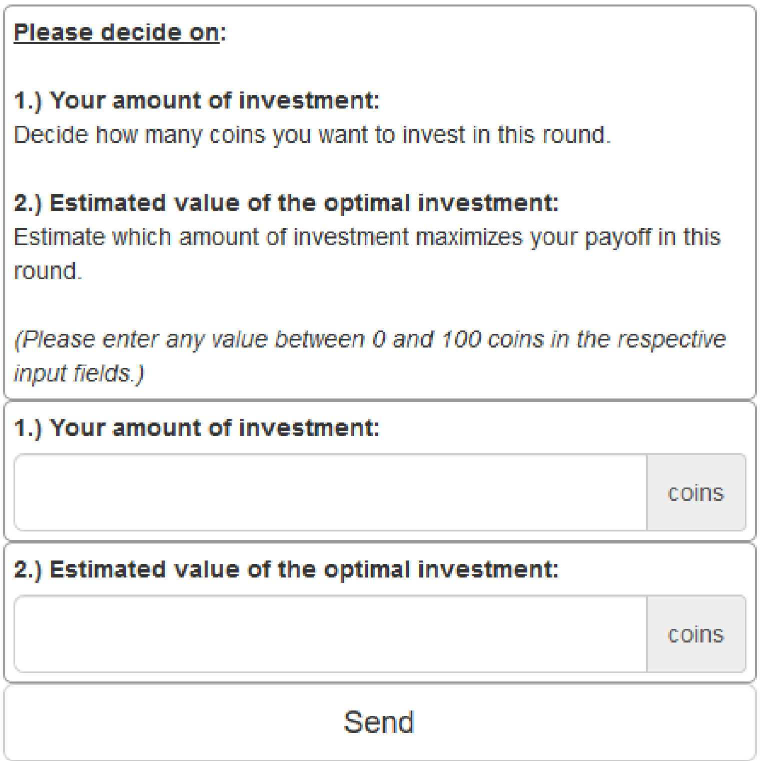


Figure B-14. Input screen for the investor – Part 2/2

B.3. Consistency between players' behavior and their self-assessment

In this appendix, I will test how consistent my interpretation of both players' strategies in the CDG is with their ex post self-evaluation of their behavior. Therefore, I will discuss the most relevant findings from my post-experimental questionnaire and relate them to both players' behavior in the game. I will begin with a short description of the most relevant items of the questionnaire. Based on that, I will analyze the consistency of the observed behavior of the advisors (*B.3.1.*) with their self-assessed preference for risk and honesty. Then, I will examine how consistent the observed behavior of the investors (*B.3.2.*) is with their self-assessed preference for risk and trust.

In my post-experimental questionnaire, I asked *all players* to rate…

…their *preference for risk* within the experiment on a 7-point-scale from completely risk-averse to completely risk-seeking.

…the *preference for risk of the other players* within the experiment on a 7-point-scale from completely risk-averse to completely risk-seeking.

Moreover, I asked the *advisors* to rate…

…their *honesty* within the experiment on a 7-point-scale from completely dishonest to completely honest.

…the *honesty of the other advisors* within the experiment on a 7-point-scale from completely dishonest to completely honest.

In addition, I asked the *investors* to rate…

…their *trust* within the experiment on a 7-point-scale from completely mistrusting to completely trusting.

…the *trust of the other investors* within the experiment on a 7-point-scale from completely mistrusting to completely trusting.

Note that in the following evaluation of the questionnaire data all 7-point-scales are coded from 0 (lowest) to 6 (highest).

B.3.1. Self-assessed lying behavior (advisors)

In the first place, I focus on the advisors' self-assessment of their *preference for risk*. This self-assessed risk preference was related to more dishonest behavior in the game, since it is significantly positively correlated with the percentage extent to which the advisors lied (L) on average over all ten rounds (Spearman's rank correlation: $\rho = 0.581$ with $p < 0.001$). In addition, their self-assessed preference for risk is significantly negatively correlated with their self-assessed honesty (Spearman's rank correlation: $\rho = -0.425$ with $p = 0.019$). This suggests that the advisors considered dishonest strategies, especially those that include lying by overstating, as more risk-seeking than honest strategies. Interestingly, they rated their own preference for risk as moderately but barely non-significantly higher than that of the other players (ratings: 3.81 vs. 3.19; two-sided Wilcoxon signed-rank test: $p = 0.096$), which indicates that they slightly overestimated their own preference for risk in relation to the group. Therefore, it makes sense to have a closer look at advisors who considered themselves as more risk-seeking than others. These advisors lied on average over all ten rounds to a significantly higher percentage extent (L) than others (193.79% vs. 119.24%; two-sided Mann-Whitney U test: $p = 0.006$). It follows that lying was associated with experiencing one's behavior as more risk-seeking than the behavior of the rest of the group.

> **Finding B-1.** *The higher the advisors' percentage extent of lying (L), the more risk-seeking they evaluated their own behavior.*

In the second place, the advisors' self-assessment of their *honesty* in the game did not differ significantly from their assessment of the honesty of the other advisors (ratings: 1.83 vs. 1.57; two-sided Wilcoxon signed-rank test: $p = 0.453$). This indicates that their self-assessed honesty was consistent in relation to the group. It comes as no surprise that the advisors' self-assessed honesty is significantly negatively correlated with the absolute value of the percentage extent to which they lied ($|L|$)[124] on average over all ten rounds (Spearman's rank correlation: $\rho = -0.542$ with $p = 0.002$). In addition, their self-assessed honesty correlates, on the one hand, significantly positively with the rate at which they engaged in cooperative truth-telling (Spearman's rank correlation: $\rho = 0.415$ with $p = 0.023$) and, on the other hand, significantly negatively with the rate at which they engaged in selfish lying (Spearman's rank correlation: $\rho = -0.536$ with $p = 0.002$). It can be concluded that the advisors' lying behavior is largely consistent with their ex post evaluation of their own honesty. In particular, the advisors considered truth-telling and cooperative behavior as honest, while considering selfish lying as dishonest, which is consistent with my taxonomy of lies and truth-telling.

> **Finding B-2.** *The advisors' lying behavior (L) and their pursued strategies based on the taxonomy of lies and truth-telling are largely consistent with the advisors' self-assessment of their honesty.*

B.3.2. Self-assessed mistrust (investors)

The investors' self-assessment of their *preference for risk* was not significantly different from their assessment of the risk preference of the other players (ratings: 3.26 vs. 3.52; two-sided Wilcoxon signed-rank test: $p = 0.350$). This indicates that their self-assessed preference for risk was consistent in relation

[124] Here, I use the absolute value of the percentage extent of lying ($|L|$), since the advisors' self-assessment of their (dis)honesty did not differentiate between lying by over- and lying by understating. However, both of these types of lies can be considered as dishonest behavior.

to the group. Moreover, the investors' self-assessed preference for risk correlates significantly positively with the percentage extent to which they engaged in mistrusting behavior ($\bar{T}$) on average over all ten rounds (Spearman's rank correlation with outlier-cleaned values: ρ = 0.552 with p = 0.002). It follows that they perceived risk-seeking mistrust in the game in fact as risk-seeking and risk-reducing mistrust as risk-averse. This means that my classification of the investors' mistrust based on its inherent risk is highly consistent with the investors' ex post evaluation of their own preference for risk in the experiment.

> **Finding B-3.** *The inherent risk of the investors' mistrusting behavior ($\bar{T}$) is consistent with their self-assessment of their preference for risk.*

Turning to the investors' self-assessment of their *trust* in the game reveals that their evaluation of their own trust barely differed from their assessment of the trust of the other investors (ratings: 2.45 vs. 2.58; two-sided Wilcoxon signed-rank test: p = 0.430). Thus, their self-assessed trust was consistent in relation to the group. In addition, the investors' self-assessed trust correlates significantly negatively with the absolute value of the percentage extent to which they engaged in mistrusting behavior ($|\bar{T}|$)[125] on average over all ten rounds (Spearman's rank correlation with outlier-cleaned values: ρ = -0.554 with p = 0.002). This indicates that the investors considered both risk-reducing and risk-seeking mistrust as mistrusting. Beyond that, their self-assessed trust is significantly positively correlated with the rate at which they engaged in trusting behavior (which corresponds to either unprofitable, cooperative, or benevolent trust) on average per trust rating (Spearman's rank correlation: ρ = 0.883 with p = 0.008). From this it follows that the investors considered trusting behavior actually as trusting, which is in line with my taxonomy of mistrust and trust. It can be concluded that the investors' mistrusting behavior is strongly consistent with their ex post evaluation of their own trust.

> **Finding B-4.** *The investors' mistrust ($\bar{T}$) and their pursued strategies based on the taxonomy of mistrust and trust are highly consistent with the investors' self-assessment of their trust.*

[125] I use the absolute value of the percentage extent of mistrust ($|\bar{T}|$) here because the investors' self-assessment of their (mis)trust did not differentiate between risk-reducing and risk-seeking mistrust. However, both of these types of mistrust can be considered as mistrusting behavior.

B.4. Temporal consistency of player decisions in the CDG

In this appendix, I test the temporal consistency of both players' behavior and first-order beliefs in the CDG. I begin with the advisors (*B.4.1.*) and then continue with the investors (*B.4.2.*).

B.4.1. Temporal consistency of advisor decisions

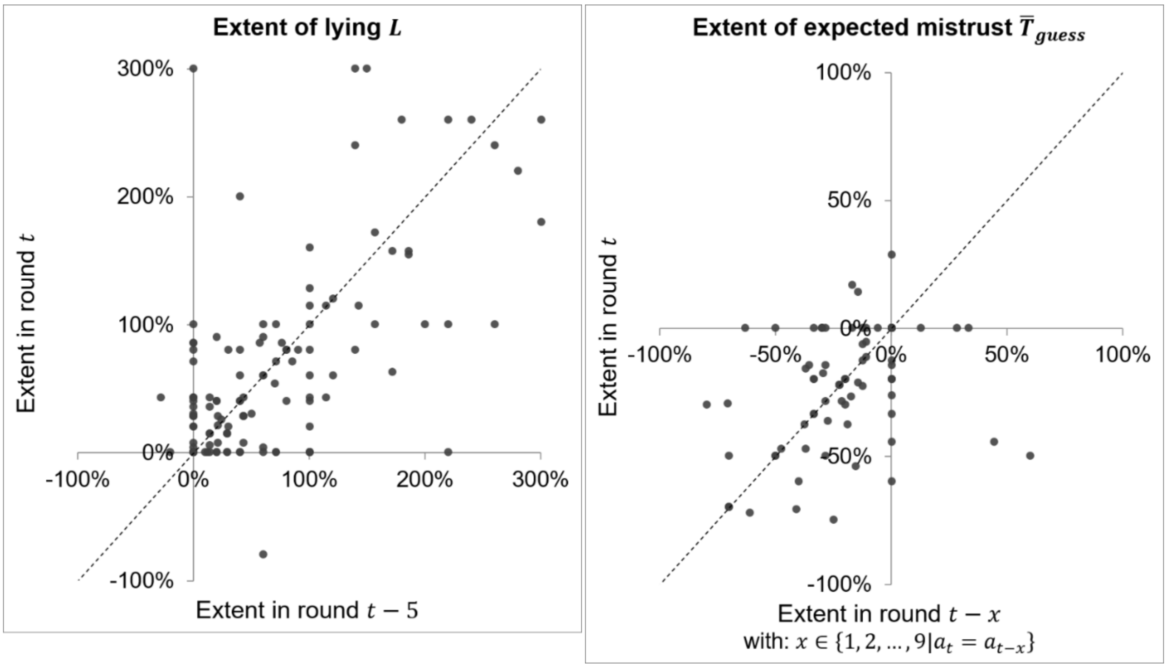
Figure B-15 visualizes lag plots for the percentage extent of lying (L) on the left (B-15a) and the percentage extent of expected mistrust ($\bar{T}_{guess}$) on the right (B-15b).[126] Note that the percentage extent of lying can be expected to depend on the value of the optimal investment. In order to analyze the temporal consistency of the advisors' lying behavior, it therefore makes sense to compare only rounds with identical optimal investments. Since each value of the optimal investment was used twice with a lag of five rounds, the values on the abscissa in the plot on the left (Figure B-15a) are lagged by five rounds. As for the plot on the right (Figure B-15b), it should be reminded that the percentage extent of expected mistrust is expected to depend on the given advice. Thus, to examine the temporal consistency of the advisors' first-order beliefs, it is reasonable to compare only rounds with identical advice. For that reason, the plot on the right (Figure B-15b) considers only advisors who gave the same advice at least twice. Here, the time lag of the values on the abscissa ranges from one to nine rounds, depending on how many rounds passed between the first and the second time that an advisor gave the same advice.[127]

The lag plot for the percentage extent of lying (L) in Figure B-15a shows that the *advisors' lying behavior* was largely consistent over time, since the percentage extent of lying correlates significantly positively with its time-lagged values (Spearman's rank correlation: $\rho = 0.538$ with $p < 0.001$). This is a result of the fact that most points in the plot are located in the first quadrant, which represents lying by overstating at both points in time. However, the plot reveals that there were different trends in the development of the advisors' (dis)honesty over time: Firstly, some advisors lied only the first time that an optimal investment was used but gave truthful advice the second time (points on the abscissa). Secondly, some advisors did the same but in reverse order (points on the ordinate). Thirdly, some advisors lied by overstating twice to the same extent when a value of the optimal investment was repeated (points on the dotted diagonal line). Finally, most of the remaining advisors also lied by overstating both times but the extent of their overstatement changed over time. Overall, in 70.32% of cases, the advisors' lying behavior had the same orientation before and after the time lag.[128]

Finding B-5. *The advisors' lying behavior (L) was mostly consistent over time.*

---

[126] For the purpose of illustration, only the most relevant section of the plot in Figure B-15a is displayed.
[127] Note here that the extent to which the advisors changed their first-order beliefs over time seems not to depend on the length of the time lag, since the number of lagged rounds is *not* significantly correlated with the change in the extent of expected mistrust over the time lag (Spearman's rank correlation: $\rho = -0.123$ with $p = 0.285$).
[128] This refers to whether they gave honest advice, lied by understating, or lied by overstating.



(a) **(b)**

Figure B-15. Temporal consistency of the advisors' behavior and first-order beliefs: (a) Lag plot of the extent of lying L; (b) Lag plot of the extent of expected mistrust $\bar{T}_{guess}$

Turning to the lag plot for the percentage extent of expected mistrust ($\bar{T}_{guess}$) in Figure B-15b, it can be seen that the *advisors' first-order beliefs* about their investors' mistrust were also generally consistent over time. In line with this, the percentage extent of expected mistrust correlates significantly positively with its time-lagged values (Spearman's rank correlation: $\rho = 0.482$ with $p < 0.001$). This is because most points in the plot are located in the third quadrant, which represents expectations of risk-reducing mistrust at both points in time. However, the plot shows several different trends in the development of the advisors' beliefs about their investors' mistrust over time: Firstly, some advisors expected mistrust from their investors the first time they gave advice and then expected trust when they gave it the second time (points on the abscissa). Secondly, some advisors had the same expectations over time but in reverse order (points on the ordinate). Thirdly, some advisors expected the same extent of mistrust from their investors in both instances when they gave the same advice twice (points on the dotted diagonal line). Fourthly, most of the remaining advisors also expected their investors to engage in risk-reducing mistrust when they gave the same advice two times but each time to a different extent. On the whole, in 68.89% of cases, the advisors did not change the orientation of their first-order beliefs over time.[129]

Finding B-6. *The advisors' first-order beliefs about their investors' mistrust ($\bar{T}_{guess}$) were mostly consistent over time.*

[129] This refers to whether they expected trusting behavior, risk-reducing mistrust, or risk-seeking mistrust from their investors.

B.4.2. Temporal consistency of investor decisions

Figure B-16 displays lag plots for the percentage extent of mistrust ($\bar{T}$) on the left (B-16a) and the percentage extent of suspected lying (L_{guess}) on the right (B-16b).[130] Here, it can be expected that both the percentage extent of mistrust and the percentage extent of suspected lying depend on the value of the received advice. Thus, to analyze the temporal consistency of the investors' behavior and first-order beliefs, it is reasonable to compare only rounds with identical advice. For that reason, both plots in Figure B-16 consider only investors who received the same advice at least twice. As a result, the time lag of the values on the abscissa ranges from one to nine rounds, depending on how many rounds passed between the first and the second time that the respective investor received advice with the same value.[131]

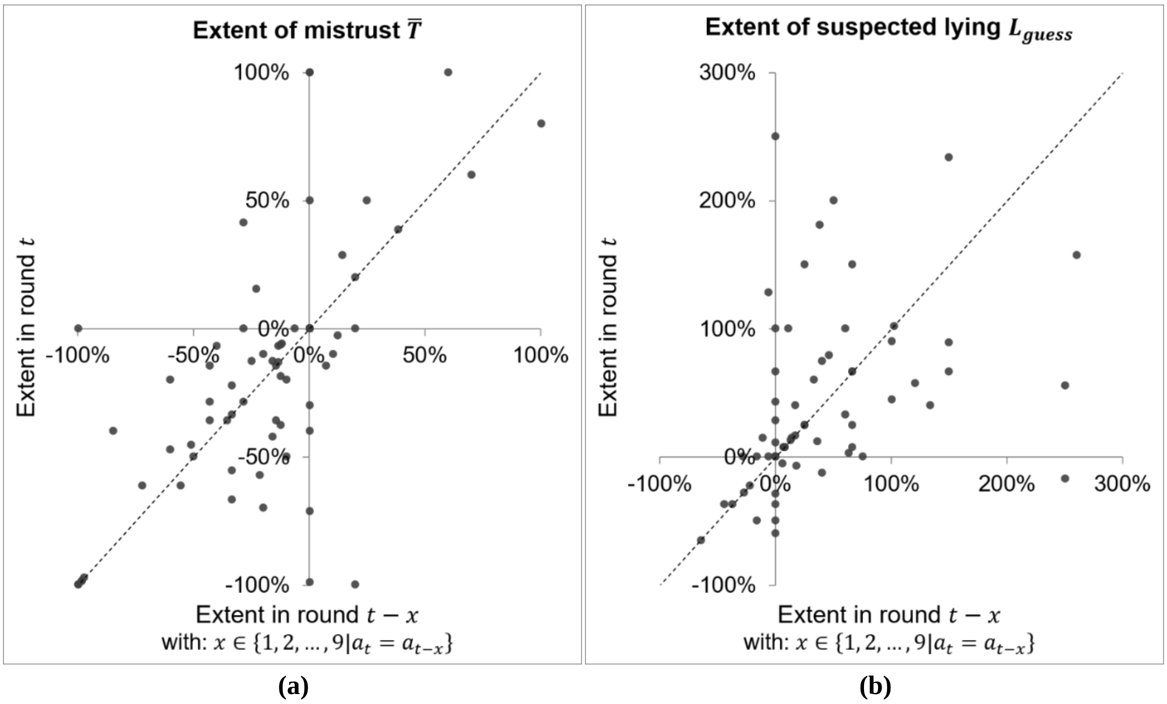


Figure B-16. Temporal consistency of the investors' behavior and first-order beliefs: (a) Lag plot of the extent of mistrust $\bar{T}$; (b) Lag plot of the extent of suspected lying L_{guess}

The lag plot for the percentage extent of mistrust ($\bar{T}$) in Figure B-16a reveals that the *investors' mistrusting behavior* was generally consistent over time, since the percentage extent of mistrust correlates significantly positively with its time-lagged values (Spearman's rank correlation: $\rho = 0.627$ with $p < 0.001$). The main reason for this is that most points in the plot are located in the third quadrant, which refers to investors who engaged in risk-reducing mistrust at both points in time. However, the

[130] For the purpose of illustration, only the most relevant section of the plot in Figure B-16b is displayed.

[131] Note that the number of lagged rounds is *neither* significantly correlated with the change in the extent of mistrust over the time lag (Spearman's rank correlation: $\rho = 0.015$ with $p = 0.905$) *nor* with the change in the extent of suspected lying over the time lag (Spearman's rank correlation: $\rho = -0.109$ with $p = 0.304$). This indicates that the extent to which the investors changed their behavior and first-order beliefs over time does not depend on the length of the time lag.

plot shows several different trends in the development of the investors' (mis)trust over time: Firstly, some investors mistrusted their received advice the first time it was given to them but trusted it the second time (points on the abscissa). Secondly, some investors did the same but in reverse order (points on the ordinate). Thirdly, some investors engaged in mistrusting behavior to the same extent in both instances when they received advice with the same value twice (points on the dotted diagonal line). Fourthly, most of the remaining investors mistrusted their advisors both times they received advice with the same value, however, each time to a different extent. Overall, in 77.63% of cases, the investors' mistrusting behavior had the same orientation before and after the time lag.[132]

Finding B-7. *The investors' mistrusting behavior ($\bar{T}$) was mostly consistent over time.*

It can be read from the lag plot of the percentage extent of suspected lying (L_{guess}) in Figure B-16b that the *investors' first-order beliefs* about their advisors' lying behavior were also generally consistent over time. This is reflected in the fact that the percentage extent of suspected lying correlates significantly positively with its time-lagged values (Spearman's rank correlation: $\rho = 0.605$ with $p < 0.001$). Moreover, it can be seen that most points in the plot are located in the first quadrant, which represents expectations of being lied to by overstating at both points in time. However, there were different trends in the development of the investors' beliefs about their advisors' lying behavior over time: Firstly, some investors suspected a piece of advice to be a lie the first time they received it but expected it to be true the second time (points on the abscissa). Secondly, some investors had the same expectations over time but in reverse order (points on the ordinate). Thirdly, some investors expected the same extent of lying from their advisors both times they received advice with the same value (points on the dotted diagonal line). Fourthly, most of the remaining investors suspected their advisors to have overstated the optimal investment both times they received advice with the same value but each time to a different extent. Overall, in 73.68% of cases, the investors did not change the orientation of their first-order beliefs over time.[133]

Finding B-8. *The investors' first-order beliefs about their advisors' lying behavior (L_{guess}) were mostly consistent over time.*

[132] This refers to whether they followed their received advice, engaged in risk-reducing mistrust, or engaged in risk-seeking mistrust.
[133] This refers to whether they expected their advisors to tell the truth, to lie by understating, or to lie by overstating.

B.5. Bootstrap methods

This appendix describes the bootstrap methods that I use in the paper in Chapter 3. All players in my experiment participated in ten rounds of the CDG. Since repeated observations between rounds are statistically dependent, certain results in Chapter 3 and its appendices are based on cluster bootstrap procedures that account for within correlations between individual decisions in different rounds. These bootstrap procedures are implemented by bootstrapping players rather than observations in the sample.[134] All bootstrap tests are performed with 100,000 bootstrap samples. Note that all bootstrap tests are two-sided.

Bootstrap unconditional proportion test. For the comparison of two proportions, I use a cluster bootstrap unconditional proportion procedure that tests whether the difference between the two proportions is different from zero based on the distribution of all bootstrap differences between both proportions. The procedure that I use refers to Efron and Tibshirani (1993).

Bootstrap binomial test. In order to test whether an observed proportion differs from a specific value (i.e., from another, given proportion with a fixed value), I use a cluster bootstrap binomial procedure that tests whether the difference between the observed and the given proportion is different from zero based on the distribution of all bootstrap differences between both proportions.

Bootstrap correlation coefficient test (for p-values of Spearman's rank correlations). In Chapter 3 and its appendices, the p-values reported for Spearman's rank correlations that are based on observations from different rounds come from cluster bootstrap procedures that test whether the respective correlation coefficient is different from zero based on the distribution of all bootstrap correlation coefficients.

Two-sample bootstrap correlation coefficient test (to compare two Spearman's rank correlations). Since my data is not normally distributed, a Fisher's Z test would be inappropriate to compare two correlation coefficients (Duncan & Layard, 1973). Therefore, and in order to account for within correlations between individual observations in different rounds, I use a cluster bootstrap procedure that tests the difference between two correlation coefficients based on the distribution of all bootstrap differences between both correlations. For independent correlation coefficients I refer to the bootstrap procedure of Rousselet et al. (2019), and for dependent correlation coefficients I refer to that of Wilcox (2016).

[134] When estimating confidence intervals for correlations of variables with their time-lagged values, the examined temporal effects are preserved by bootstrapping pairs of observations that come from the same player but from two different time points (i.e., rounds) with the respective time lag in the sample.

C. Appendices to Chapter 4

This section consists of four separate appendices to Chapter 4. The first appendix specifies the experimental design by providing both the exact definition and an example of both players' payoff functions (*C.1.*). The second appendix provides information on the instructions that I presented to the subjects in the experiment (*C.2.*). The third appendix presents additional results (*C.3.*). Finally, the fourth appendix provides details on the bootstrap procedures that I use in the paper in Chapter 4 (*C.4.*).

C.1. Payoff structure

Both players' payoff functions are defined in line with Beck's (2020) Continuous Deception Game (CDG), which is described in full detail in Chapter 3. On the one hand, the payoff of the sender $\pi_S(I)$ is defined as:

$$\pi_S(I) = 0.5 * I \tag{66}$$

with I referring to the size of the investment made by the receiver. As can be seen, the payoff of the sender increases linearly with the size of the investment.

On the other hand, the payoff of the receiver $\pi_R(I)$ is defined as:

$$\pi_R(I) = \begin{cases} I \leq I^*: & 100 + 0.5 * I \\ I > I^*: & 100 + I^* - 0.5 * I \end{cases} \tag{67}$$

with I referring to the size of the investment and I^* being the true value of the optimal investment. This payoff function ensures that the receiver's payoff decreases by any upward or downward deviation of the investment from the true optimal investment. Note that, in the experiment, I use coins as an in-game currency (Beck, 2020). The subjects can exchange these coins to euros at the end of the experiment. For 100 coins they receive 8 euros, whereby their off-game payoffs are rounded up to the nearest 10 cent value.

Figure C-1 shows an example of both players' payoff functions. Here, the optimal investment is assumed to be 50 coins. To visualize the payoff structure to the players, in each round of the CDG a similar figure with the current round's optimal investment is presented to the sender. Since this figure reveals the location of the optimal investment, the receiver is shown a different animated version of this figure, in which the location of the optimal investment and the corresponding peak of their payoff function is moving across the screen from left to right and back again (Beck, 2020). More details on both players' instructions can be found in appendix C.2.

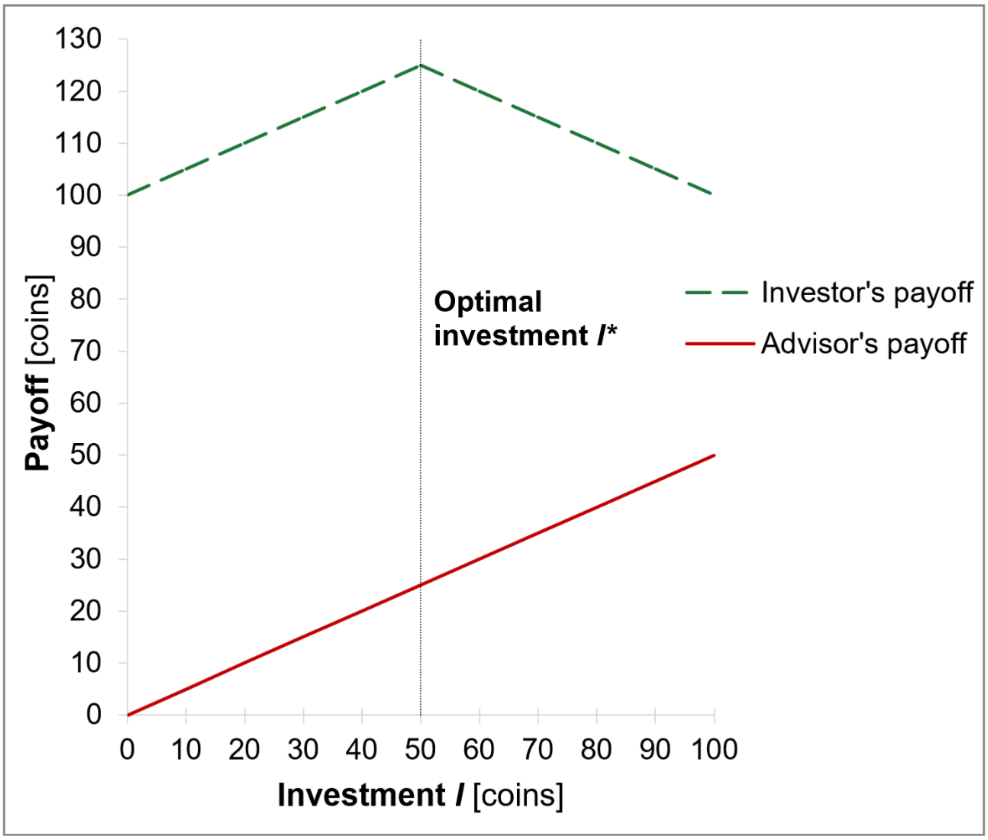


Figure C-1. Payoff structure of the CDG [revised version of original figure from Beck (2020)]

C.2. Instructions

C.2.1. Baseline treatment

The baseline treatment in the paper in Chapter 4 is identical to Beck's (2020) CDG (on which I report in Chapter 3). For this reason, the instructions that I use for the baseline treatment in Chapter 4 are identical to the instructions for the experiment in Chapter 3 and can be found in appendix B.2.

C.2.2. Honesty oath treatment

Figure C-2 shows a translation of the honesty oath that the subjects signed in the honesty oath treatment in the paper in Chapter 4.[135] Original materials in German are available upon request.

Oath

"Hereby I do swear that my actions during this experiment will be due to the

principle of honesty**.**

In particular, I swear not to lie in order to enrich myself."

Place and date

Signature

Figure C-2. Honesty oath

[135] Supplementary instructions regarding the experimental procedure (including information on the player IDs, technical instructions on the use of *classEx* and the confirmation of payments) are omitted here.

C.3. Additional results

This appendix presents the following additional results for the paper in Chapter 4: Firstly, an overview of the average extent of lying in all rounds and its differences between both treatments (*C.3.1.*). Secondly, an overview of the average extent to which liars lied in all rounds and its differences between both treatments (*C.3.2.*). Thirdly, a comparison of the distribution and the standard deviation of the extent to which liars lied in all rounds between both treatments (*C.3.3.*). Fourthly, an overview of the average decision time in all rounds and the testing of decision time differences between both treatments as well as between liars and truth-tellers (*C.3.4.*). Fifthly, the testing of truth ratio differences between the decision time quarters in both treatments (*C.3.5.*). Sixthly, additional results on the temporal consistency of the senders' truth-telling behavior (*C.3.6.*). Lastly, an analysis of the receivers' first-order beliefs (*C.3.7.*).

C.3.1. Average extent of lying in all rounds of both treatments

Round	Baseline treatment	Honesty oath treatment	Difference of averages (1^{st}-2^{nd})	p-value (two-sided Mann-Whitney U test[1])
1	38.06%	8.52%	29.54%	0.002
2	11.52%	4.47%	7.05%	0.116
3	127.74%	69.68%	58.06%	0.010
4	76.19%	40.09%	36.10%	0.013
5	459.35%	317.74%	141.61%	0.067
6	39.29%	10.97%	28.32%	0.002
7	18.80%	7.37%	11.43%	0.009
8	122.06%	69.94%	52.12%	0.012
9	68.94%	41.90%	27.04%	0.044
10	519.00%	257.33%	261.67%	0.002
Average	148.10%	82.80%	65.30%	0.004

[1] Mann-Whitney U test to compare the mean rank of the extent of lying between both treatments.

Table C-1. Average extent of lying (all senders)

C.3.2. Average extent to which liars lied in all rounds of both treatments

Round	Baseline treatment	Honesty oath treatment	Difference of averages (1^{st}-2^{nd})	p-value (two-sided Mann-Whitney U test[1])
1	51.30%	17.60%	33.70%	0.106
2	21.01%	9.90%	11.11%	0.187
3	141.43%	135.00%	6.43%	0.864
4	98.41%	77.68%	20.73%	0.438
5	508.57%	518.42%	-9.85%	0.870
6	52.96%	28.33%	24.63%	0.172
7	26.49%	19.05%	7.44%	0.898
8	145.54%	144.53%	1.01%	0.892
9	85.49%	72.17%	13.32%	0.843
10	595.89%	498.58%	97.31%	0.338

[1] Mann-Whitney U test to compare the mean rank of the extent of lying between both treatments.

Table C-2. Average extent of lying (only liars)

C.3.3. Standard deviation of the extent to which liars lied in all rounds of both treatments

Round	Baseline treatment	Honesty oath treatment	Difference of standard deviations (1st-2nd)	p-value (non-parametric Levene's test[1])	p-value (Kolmogorov-Smirnov test[2])
1	26.68%	55.13%	-28.45%	0.964	0.265
2	16.56%	24.32%	-7.76%	0.716	0.807
3	91.40%	88.99%	2.41%	0.855	0.955
4	52.57%	69.59%	-17.02%	0.570	0.482
5	336.63%	335.50%	1.13%	0.692	0.968
6	40.24%	54.08%	-13.84%	0.917	0.284
7	12.50%	27.97%	-15.47%	0.570	0.982
8	96.85%	118.71%	-21.86%	0.623	0.996
9	46.38%	76.74%	-30.36%	0.264	0.680
10	260.41%	314.03%	-53.62%	0.405	0.605

[1] Non-parametric Levene's test to compare the variance of the extent of lying between both treatments; [2] Kolmogorov-Smirnov test to compare the distribution of the extent of lying between both treatments.

Table C-3. Standard deviation of the extent of lying (only liars)

C.3.4. Average decision time in all rounds of both treatments

Round	Baseline treatment	Honesty oath treatment	Difference of averages (1st-2nd)	p-value (two-sided Mann-Whitney U test[1])
1	59.55	59.99	-0.44	0.894
2	46.22	45.26	0.96	0.905
3	39.15	46.54	-7.39	0.073
4	35.63	39.79	-4.16	0.379
5	38.56	39.37	-0.81	0.426
6	30.52	28.23	2.29	0.688
7	23.98	26.65	-2.67	0.569
8	23.49	25.29	-1.80	0.435
9	21.71	27.49	-5.78	0.229
10	25.32	23.12	2.20	0.418
Average	34.41	36.17	-1.76	0.550

[1] Mann-Whitney U test to compare the mean rank of the decision time between both treatments.

Table C-4. Average decision time

Round	p-values of two-sided Mann-Whitney U tests[1]			
	BT: *Lying vs. Truth-telling*	HOT: *Lying vs. Truth-telling*	Lying: *BT vs. HOT*	Truth-telling: *BT vs. HOT*
1	0.928	0.304	0.411	0.742
2	0.525	0.500	0.905	0.937
3	0.385	0.252	0.060	0.824
4	0.571	0.033	0.068	1.000
5	0.640	0.068	0.129	0.365
6	0.652	0.004	0.144	0.515
7	0.514	0.003	0.009	0.142
8	0.591	0.002	0.013	0.398
9	0.271	<0.001	0.005	0.831
10	0.316	<0.001	0.248	0.411

[1] Mann-Whitney U test to compare the mean rank of the decision time between the respective groups.

Table C-5. Testing of decision time differences

C.3.5. Truth ratio differences between decision time quarters

	p-values of bootstrap unconditional proportion tests[1]			
Decision time	*Very fast*	*Fast*	*Slow*	*Very slow*
Very fast	-	0.624	0.770	0.572
Fast	-	-	0.809	0.233
Slow	-	-	-	0.223
Very slow	-	-	-	-

[1] Bootstrap unconditional proportion test to compare the overall truth ratio between the respective decision time quarters.

Table C-6. Testing of truth ratio differences between decision time quarters in the baseline treatment

	p-values of bootstrap unconditional proportion tests[1]			
Decision time	*Very fast*	*Fast*	*Slow*	*Very slow*
Very fast	-	0.086	<0.001	<0.001
Fast	-	-	0.005	<0.001
Slow	-	-	-	0.184
Very slow	-	-	-	-

[1] Bootstrap unconditional proportion test to compare the overall truth ratio between the respective decision time quarters.

Table C-7. Testing of truth ratio differences between decision time quarters in the honesty oath treatment

C.3.6. Temporal consistency of truth-telling behavior

In the results section of the paper in Chapter 4, I report that the proportion of senders who never lied throughout the entire experiment is significantly higher with an honesty oath than without it (BT: 0% vs. HOT: 29%; two-sided Fisher exact test: p = 0.002). In this appendix, I will show that this difference between both treatments cannot be explained solely by the fact that more senders told the truth under oath. This will allow me to draw conclusions on whether the honesty oaths made the senders' truth-telling behavior more consistent over time.

Since repeated observations are statistically dependent, the truth-telling behavior of the senders is expected to be at least partially consistent over time in both treatments. Moreover, if senders who never lied in the experiment are overrepresented, it can be assumed that some of the senders based their truth-telling behavior in later rounds on their decision to tell the truth in previous rounds.[136] There are two plausible explanations of why senders who never lied in all ten rounds could be overrepresented: On the one hand, the preference for honesty of some players might be so high that these players simply do not want to lie in the experiment. If this was correct, senders who never lied would be expected to be overrepresented in both treatments. On the other hand, the honesty oath could have made the senders'

[136] Note that "overrepresented" in this context means that disproportionally more senders told the truth in all ten rounds than it would be expected if their decisions to tell the truth were independent between rounds. Here the expected proportion of senders who never lied in a treatment depends on the overall truth ratio in this treatment.

truth-telling behavior more consistent over time. In this case, senders who never lied would be expected to be overrepresented only in the honesty oath treatment.

In order to test whether senders who never lied are overrepresented in a treatment, I compare the observed probability that a random sender never lied in this treatment to its expected value. Here, the expected value is based on the observed truth ratio in the respective treatment and calculated under the assumption that the senders' truth-telling behavior is not dependent between rounds.

In the baseline treatment, the observed truth ratio was 22%. Thus, under the assumption of independent observations between rounds, the probability of a sender telling the truth in all ten rounds is close to 0% (= 0.22^{10} = 0.00002%). Here, a one-sided binomial test that tests whether the observed proportion of senders who never lied without an honesty oath (0%) is higher than its expected value (0.00002%) is not significant ($p = 0.969$). On this basis, the null hypothesis that senders who told the truth in all ten rounds are *not* overrepresented in the baseline treatment cannot be rejected. Thus, without the honesty oath, I do not find that the senders' truth-telling behavior was more consistent over time than one would expect it to be if their decisions to tell the truth were independent between rounds.

By contrast, in the honesty oath treatment, the senders told the truth in about 51% of cases. Again, under the assumption of independent observations between rounds, the probability that a sender tells the truth in all ten rounds after taking an honesty oath is about 0.1% (= 0.51^{10}). In this treatment, the proportion of senders who never lied (29%) is significantly higher than its expected value (0.1%), according to a one-sided binomial test ($p < 0.001$). Hence, senders who never lied throughout the experiment are overrepresented in the honesty oath treatment, which suggests that some senders based their truth-telling behavior in later rounds on their decision to tell the truth in previous rounds.

Since the conducted binomial tests already consider the different truth ratios in both treatments, the higher proportion of senders who never lied in the honesty oath treatment cannot be explained by the fact that more senders told the truth under oath. It can be concluded that the senders' decision to tell the truth was more consistent over time with an honesty oath than without it.

C.3.7. First-order beliefs of the receivers

Figure C-3 displays the overall average extent to which the receivers suspected the senders to lie (over all rounds and receivers) in both treatments. As can be read from the figure, this extent is significantly lower when the receivers were informed about the honesty oath than when they were not (BT: 56% vs. HOT: 19%; two-sided Mann-Whitney U test: $p = 0.013$). This result holds true when the comparison is based on pooled data of all rounds (Bootstrap mean test: $p = 0.002$).

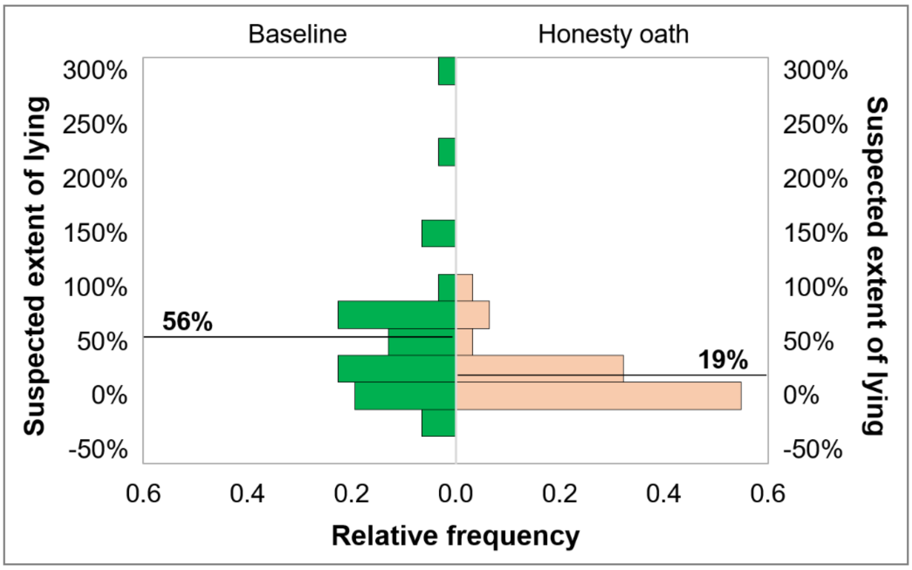


Figure C-3. Distribution of the suspected extent of lying

Moreover, in the honesty oath treatment, in 46% of observed decisions the receivers expected the senders to tell the truth, which is an increase of about 109% in comparison to the baseline treatment. This increase in the overall expected truth ratio (over all rounds and receivers) is significant (BT: 22% vs. HOT: 46%; bootstrap unconditional proportion test: $p < 0.001$). Finally, when only the receivers who suspected the senders to lie are considered, the average suspected extent of lying is significantly lower in the honesty oath treatment than in the baseline treatment (Bootstrap mean test on pooled data of all rounds: $p = 0.028$). It can be concluded that the receivers expected the honesty oath both to increase the overall truth ratio among the senders and to reduce the extent to which liars lie. While the former expected effect of the oath is consistent with the actual change in the senders' lying behavior, the latter is not. This indicates that the honesty oath disproportionally fostered the receivers' trust in the senders.

Finding C-1. *Informing subjects that other players took an honesty oath increases their trust in the honesty of the oath-takers.*

C.4. Bootstrap methods

This appendix describes the bootstrap methods that I use in the paper in Chapter 4.[137] All players in my experiment participated in ten rounds of Beck's (2020) CDG. Since repeated observations between rounds are correlated, certain results in Chapter 4 and its appendices are based on cluster bootstrap procedures that account for within correlations between individual decisions in different rounds. These bootstrap procedures are implemented by bootstrapping players rather than observations in the sample.[138] All bootstrap tests are performed with 100,000 bootstrap samples.

Bootstrap mean test. For the comparison of two means, I use a cluster bootstrap procedure that tests whether the difference between the two means is different from zero based on the distribution of all bootstrap differences between both means. The procedure that I use refers to Efron and Tibshirani (1993). All respective tests are two-sided.

Bootstrap unconditional proportion test. For the comparison of two proportions, I use a cluster bootstrap unconditional proportion procedure that tests whether the difference between the two proportions is different from zero based on the distribution of all bootstrap differences between both proportions. This bootstrap procedure is performed analogously to the bootstrap mean test. All respective tests are two-sided.

Bootstrap correlation coefficient test (for p-values of Spearman's rank correlations). The p-values reported for all Spearman's rank correlations in Chapter 4 and its appendices come from cluster bootstrap procedures that test whether the respective correlation coefficient is different from zero based on the distribution of all bootstrap correlation coefficients. All respective tests are two-sided.

Two-sample bootstrap correlation coefficient test (to compare two Spearman's rank correlations). Since my data is not normally distributed, a Fisher's Z test would be inappropriate to compare two correlation coefficients (Duncan & Layard, 1973). Therefore, and in order to account for within correlations between individual observations in different rounds, I use a cluster bootstrap procedure that tests the difference between two correlation coefficients based on the distribution of all bootstrap differences between both correlations. For independent correlation coefficients I refer to the bootstrap procedure of Rousselet et al. (2019), and for dependent correlation coefficients I refer to that of Wilcox (2016). In the paper in Chapter 4, all respective tests are two-sided.

---

[137] Note that there are small differences between the bootstrap methods that I use in Chapter 3 and Chapter 4. To account for these differences, each of these chapters has a separate appendix explaining the corresponding bootstrap methods. For this reason, the appendix on bootstrap methods to Chapter 4 is similar to (but slightly different from) that to Chapter 3.

[138] Note that when estimating confidence intervals for autocorrelations of a given time lag, the examined temporal effects are preserved by bootstrapping pairs of observations that come from the same player but from two different time points (i.e., rounds) with the respective time lag in the sample.

Bibliography

Abbink, K. (2000). Fair salaries and the moral costs of corruption. *Bonn Econ Discussion Papers, 1/2000*.

Abeler, J., Nosenzo, D., & Raymond, C. (2019). Preferences for Truth-Telling. *Econometrica, 87*(4), 1115-1153. https://doi.org/10.3982/ecta14673

Alger, I., & Weibull, J. W. (2013). Homo Moralis--Preference Evolution Under Incomplete Information and Assortative Matching. *Econometrica, 81*(6), 2269-2302. https://doi.org/10.3982/ecta10637

Anderson, M., & Escher, P. (2010). *The MBA Oath: Setting a Higher Standard for Business Leaders*. Penguin Group.

Ariely, D. (2012). *The Honest Truth about Dishonesty: How We Lie to Everyone--Especially Ourselves*. HarperCollins.

Armantier, O., & Boly, A. (2008). Can Corruption be Studied in the Lab? Comparing a Field and a Lab Experiment. *CIRANO – Scientific Publications, 2008s-26*. https://doi.org/10.2139/ssrn.1324120

Ashforth, B. E., & Anand, V. (2003). The Normalization of Corruption in Organizations. *Research in Organizational Behavior, 25*, 1-52. https://doi.org/10.1016/s0191-3085(03)25001-2

Bachvarova, M. R. (2008). Oath and Allusion in Alcaeus fr. 129. In A. Sommerstein & J. Fletcher (Eds.), *Horkos: The Oath in Greek Society* (pp. 179-188). Phoenix Press. https://doi.org/10.5949/liverpool/9781904675679.003.0016

Banfield, E. C. (1975). Corruption as a Feature of Governmental Organization. *The Journal of Law and Economics, 18*(3), 587-605. https://doi.org/10.1086/466826

Barefoot, J. C., Maynard, K. E., Beckham, J. C., Brummett, B. H., Hooker, K., & Siegler, I. C. (1998). Trust, Health, and Longevity. *Journal of Behavioral Medicine, 21*(6), 517-526. https://doi.org/10.1023/a:1018792528008

Barr, A., & Michailidou, G. (2017). Complicity without connection or communication. *Journal of Economic Behavior & Organization, 142*, 1-10. https://doi.org/10.1016/j.jebo.2017.07.013

Bateson, M., Nettle, D., & Roberts, G. (2006). Cues of being watched enhance cooperation in a real-world setting. *Biology Letters, 2*(3), 412-414. https://doi.org/10.1098/rsbl.2006.0509

Batson, C. D., Thompson, E. R., Seuferling, G., Whitney, H., & Strongman, J. A. (1999). Moral hypocrisy: Appearing moral to oneself without being so. *Journal of Personality and Social Psychology, 77*(3), 525-537. https://doi.org/10.1037/0022-3514.77.3.525

Battigalli, P., Charness, G., & Dufwenberg, M. (2013). Deception: The role of guilt. *Journal of Economic Behavior & Organization, 93*, 227-232. https://doi.org/10.1016/j.jebo.2013.03.033

Beck, T. (2020). Size matters! Lying and Mistrust in the Continuous Deception Game. *Verein für Socialpolitik Annual Conference 2020: Gender Economics*, ZBW - Leibniz Information Centre for Economics, Kiel, Germany. http://hdl.handle.net/10419/224530

Beck, T. (2021). How the honesty oath works: Quick, intuitive truth telling under oath. *Journal of Behavioral and Experimental Economics*, Forthcoming. https://doi.org/10.1016/j.socec.2021.101728

Beck, T., Bühren, C., Frank, B., & Khachatryan, E. (2020). Can Honesty Oaths, Peer Interaction, or Monitoring Mitigate Lying? *Journal of Business Ethics, 163*(3), 467-484. https://doi.org/10.1007/s10551-018-4030-z

Belfast Telegraph. (2011). *Bank fined £10.5m over mis-selling*. Belfast Telegraph. Retrieved 15 October 2020 from https://www.belfasttelegraph.co.uk/news/uk/bank-fined-105m-over-mis-selling-28688514.html

Bentham, J. (1817/2005). *Swear Not at All. Containing an Exposure of the Mischievousness as well as Antichristianity of the Ceremony of an Oath*. Elibron Classics.

Berg, J., Dickhaut, J., & McCabe, K. (1995). Trust, Reciprocity, and Social History. *Games and Economic Behavior, 10*(1), 122-142. https://doi.org/10.1006/game.1995.1027

Beugelsdijk, S., de Groot, H. L. F., & van Schaik, A. B. T. M. (2004). Trust and Economic Growth: A Robustness Analysis. *Oxford Economic Papers, 56*(1), 118-134. https://doi.org/10.1093/oep/56.1.118

Bjørnskov, C. (2012). How Does Social Trust Affect Economic Growth? *Southern Economic Journal, 78*(4), 1346-1368. https://doi.org/10.4284/0038-4038-78.4.1346

Blok, V. (2013). The Power of Speech Acts: Reflections on a Performative Concept of Ethical Oaths in Economics and Business. *Review of Social Economy, 71*(2), 187-208. https://doi.org/10.1080/00346764.2013.799965

Boatright, J. R. (2013). Swearing to be Virtuous: The Prospects of a Banker's Oath. *Review of Social Economy, 71*(2), 140-165. https://doi.org/10.1080/00346764.2013.800305

Bollnow, O. F. (1947). Wahrhaftigkeit. *Die Sammlung, 2*(5-6), 234-245.

Brown, M. E., & Treviño, L. K. (2006). Ethical leadership: A review and future directions. *The Leadership Quarterly, 17*(6), 595-616. https://doi.org/10.1016/j.leaqua.2006.10.004

Bucciol, A., & Piovesan, M. (2011). Luck or cheating? A field experiment on honesty with children. *Journal of Economic Psychology, 32*(1), 73-78. https://doi.org/10.1016/j.joep.2010.12.001

Bussmann, K.-D., Krieg, O., Nestler, C., Salvenmoser, S., Schroth, A., Theile, A., & Trunk, D. (2009). *Wirtschaftskriminalität 2009: Sicherheitslage in deutschen Großunternehmen*. PricewaterhouseCoopers.

Cabrera, A. (2003). *A Hippocratic Oath for business?* Handelsblatt. Retrieved 15 October 2020 from https://www.handelsblatt.com/unternehmen/management/guest-article-a-hippocratic-oath-for-business/2259900.html

Cagala, T., Glogowsky, U., & Rincke, J. (2014). A field experiment on intertemporal enforcement spillovers. *Economics Letters, 125*(2), 171-174. https://doi.org/10.1016/j.econlet.2014.08.034

Camerer, C. F., & Hogarth, R. M. (1999). The Effects of Financial Incentives in Experiments: A Review and Capital-Labor-Production Framework. *Journal of Risk and Uncertainty, 19*(1-3), 7-42. https://doi.org/10.1023/a:1007850605129

Capraro, V., Schulz, J., & Rand, D. G. (2019). Time pressure and honesty in a deception game. *Journal of Behavioral and Experimental Economics, 79*, 93-99. https://doi.org/10.1016/j.socec.2019.01.007

Carlsson, F., Kataria, M., Krupnick, A., Lampi, E., Löfgren, Å., Qin, P., & Sterner, T. (2013). The truth, the whole truth, and nothing but the truth--A multiple country test of an oath script. *Journal of Economic Behavior & Organization, 89*, 105-121. https://doi.org/10.1016/j.jebo.2013.02.003

Carson, T. L. (1988). On the definition of lying: A reply to Jones and revisions. *Journal of Business Ethics, 7*(7), 509-514. https://doi.org/10.1007/bf00382597

Carson, T. L. (2010). *Lying and Deception: Theory and Practice*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199577415.001.0001

Charness, G., & Dufwenberg, M. (2006). Promises and Partnership. *Econometrica, 74*(6), 1579-1601. https://doi.org/10.1111/j.1468-0262.2006.00719.x

Charness, G., & Dufwenberg, M. (2010). Bare promises: An experiment. *Economics Letters, 107*(2), 281-283. https://doi.org/10.1016/j.econlet.2010.02.009

Charness, G., & Sutter, M. (2012). Groups Make Better Self-Interested Decisions. *Journal of Economic Perspectives, 26*(3), 157-176. https://doi.org/10.1257/jep.26.3.157

Christens, S., Dannenberg, A., & Sachs, F. (2019). Identification of individuals and groups in a public goods experiment. *Journal of Behavioral and Experimental Economics, 82*(101445). https://doi.org/10.1016/j.socec.2019.06.010

Cleek, M. A., & Leonard, S. L. (1998). Can corporate codes of ethics influence behavior? *Journal of Business Ethics, 17*(6), 619-630. https://doi.org/10.1023/a:1017969921581

Cohen, T. R., Gunia, B. C., Kim-Jun, S. Y., & Murnighan, J. K. (2009). Do groups lie more than individuals? Honesty and deception as a function of strategic self-interest. *Journal of Experimental Social Psychology, 45*(6), 1321-1324. https://doi.org/10.1016/j.jesp.2009.08.007

Columbia Business School. (2020). *Honor Code of Columbia Business School*. Columbia Business School. Retrieved 15 October 2020 from https://www8.gsb.columbia.edu/honor/

Conrads, J., Irlenbusch, B., Rilke, R. M., & Walkowitz, G. (2013). Lying and team incentives. *Journal of Economic Psychology, 34*, 1-7. https://doi.org/10.1016/j.joep.2012.10.011

Court Organization and Civil Code: Swearing of witnesses by foreperson, NJ Rev Stat § 2B:21-6 (2013).

Cox, J. C. (2004). How to identify trust and reciprocity. *Games and Economic Behavior, 46*(2), 260-281. https://doi.org/10.1016/s0899-8256(03)00119-2

Dana, J., Weber, R. A., & Kuang, J. X. (2007). Exploiting moral wiggle room: experiments demonstrating an illusory preference for fairness. *Economic Theory, 33*(1), 67-80. https://doi.org/10.1007/s00199-006-0153-z

Danilov, A., Biemann, T., Kring, T., & Sliwka, D. (2013). The dark side of team incentives: Experimental evidence on advice quality from financial service professionals. *Journal of Economic Behavior & Organization, 93*, 266-272. https://doi.org/10.1016/j.jebo.2013.03.012

Dannenberg, A., & Khachatryan, E. (2020). A comparison of individual and group behavior in a competition with cheating opportunities. *Journal of Economic Behavior & Organization, 177*, 533-547. https://doi.org/10.1016/j.jebo.2020.06.028

Davidson, R., & MacKinnon, J. G. (1996). The Power of Bootstrap Tests. *Queen's Institute for Economic Research Discussion Papers, 937*. https://doi.org/10.22004/ag.econ.273372

de Bruin, B. (2016). Pledging Integrity: Oaths as Forms of Business Ethics Management. *Journal of Business Ethics, 136*(1), 23-42. https://doi.org/10.1007/s10551-014-2504-1

DeMartino, G. F. (2011). *The Economist's Oath: On the Need for and Content of Professional Economic Ethics*. Oxford University Press.

Dhanarajan, S. (2005). Managing ethical standards: When rhetoric meets reality. *Development in Practice, 15*(3-4), 529-538. https://doi.org/10.1080/09614520500075946

Dimmock, S. G., & Gerken, W. C. (2012). Predicting fraud by investment managers. *Journal of Financial Economics, 105*(1), 153-173. https://doi.org/10.1016/j.jfineco.2012.01.002

Dreber, A., & Johannesson, M. (2008). Gender differences in deception. *Economics Letters, 99*(1), 197-199. https://doi.org/10.1016/j.econlet.2007.06.027

Duncan, G. T., & Layard, M. W. J. (1973). A Monte-Carlo study of asymptotically robust tests for correlation coefficients. *Biometrika, 60*(3), 551-558. https://doi.org/10.1093/biomet/60.3.551

Efron, B., & Tibshirani, R. J. (1993). Hypothesis testing with the bootstrap. In B. Efron & R. J. Tibshirani (Eds.), *An Introduction to the Bootstrap* (1 ed., pp. 220-236). Chapman and Hall.

Egan, M. (2016). *5,300 Wells Fargo employees fired over 2 million phony accounts*. CNN Money. Retrieved 15 October 2020 from http://money.cnn.com/2016/09/08/investing/wells-fargo-created-phony-accounts-bank-fees/index.html

Engelmann, J. B., & Fehr, E. (2016). The slippery slope of dishonesty. *Nature Neuroscience, 19*(12), 1543-1544. https://doi.org/10.1038/nn.4441

Erat, S., & Gneezy, U. (2012). White Lies. *Management Science, 58*(4), 723-733. https://doi.org/10.1287/mnsc.1110.1449

Evans, A. M., Dillon, K. D., & Rand, D. G. (2015). Fast but not intuitive, slow but not reflective: Decision conflict drives reaction times in social dilemmas. *Journal of Experimental Psychology, 144*(5), 951-966. https://doi.org/10.1037/xge0000107

Fallis, D. (2010). Lying and Deception. *Philosophers' Imprint, 10*(12), 1-22.

Fehr, E., & List, J. A. (2004). The Hidden Costs and Returns of Incentives--Trust and Trustworthiness among Ceos. *Journal of the European Economic Association, 2*(5), 743-771. https://doi.org/10.1162/1542476042782297

Fehr, E., & Schmidt, K. M. (1999). A Theory of Fairness, Competition, and Cooperation. *The Quarterly Journal of Economics*, *114*(3), 817-868. https://doi.org/10.1162/003355399556151

Fischbacher, U., & Föllmi-Heusi, F. (2013). Lies in Disguise--An Experimental Study on Cheating. *Journal of the European Economic Association, 11*(3), 525-547. https://doi.org/10.1111/jeea.12014

Fowler, B., Franklin, C., & Hyde, R. (2001). Internet Securities Fraud: Old Trick, New Medium. *Duke Law & Technology Review, 1*(1).

Frank, B., Li, S., Bühren, C., & Qin, H. (2015). Group Decision Making in a Corruption Experiment: China and Germany Compared. *Jahrbücher für Nationalökonomie und Statistik, 235*(2), 207-227. https://doi.org/10.1515/jbnst-2015-0207

Garbarino, E., Slonim, R., & Villeval, M. C. (2017). Loss Aversion and Lying Behavior: Theory, Estimation and Empirical Evidence. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.2875989

Gaventa, J., & McGee, R. (2013). The Impact of Transparency and Accountability Initiatives. *Development Policy Review, 31*(s1), s3-s28. https://doi.org/10.1111/dpr.12017

Gerlach, P., Teodorescu, K., & Hertwig, R. (2019). The Truth About Lies: A Meta-Analysis on Dishonest Behavior. *Psychological Bulletin, 145*(1), 1-44. https://doi.org/10.1037/bul0000174

Giamattei, M., & Lambsdorff, J. G. (2019). classEx--an online tool for lab-in-the-field experiments with smartphones. *Journal of Behavioral and Experimental Finance, 22*, 223-231. https://doi.org/10.1016/j.jbef.2019.04.008

Gino, F., Ayal, S., & Ariely, D. (2009a). Contagion and differentiation in unethical behavior: the effect of one bad apple on the barrel. *Psychological Science, 20*(3), 393-398. https://doi.org/10.1111/j.1467-9280.2009.02306.x

Gino, F., Ayal, S., & Ariely, D. (2009b). Out of Sight, Ethically Fine? The Effects of Collaborative Work on Individuals' Dishonesty. *INGroup, Fourth Annual Conference, Colorado Springs*.

Gino, F., Ayal, S., & Ariely, D. (2013). Self-Serving Altruism? The Lure of Unethical Actions that Benefit Others. *Journal of Economic Behavior & Organization, 93*, 285-292. https://doi.org/10.1016/j.jebo.2013.04.005

Glöckner, A. (2009). Investigating intuitive and deliberate processes statistically: The multiple-measure maximum likelihood strategy classification method. *Judgment and Decision Making, 4*(3), 186-199.

Gneezy, U. (2005). Deception: The Role of Consequences. *American Economic Review, 95*(1), 384-394. https://doi.org/10.1257/0002828053828662

Gneezy, U., Kajackaite, A., & Sobel, J. (2018). Lying Aversion and the Size of the Lie. *American Economic Review, 108*(2), 419-453. https://doi.org/10.1257/aer.20161553

Gneezy, U., Saccardo, S., Serra-Garcia, M., & van Veldhuizen, R. (2020). Bribing the Self. *CESifo Working Papers, 8065.*

Gurtman, M. B. (1992). Trust, distrust, and interpersonal problems: A circumplex analysis. *Journal of Personality and Social Psychology, 62*(6), 989-1002. https://doi.org/10.1037/0022-3514.62.6.989

Haas, M. (2016). *Wie Trump das moralische Fundament der USA beschädigt*. Süddeutsche Zeitung. Retrieved 15 October 2020 from https://sz-magazin.sueddeutsche.de/wild-wild-west-amerikakolumne/wie-trump-das-moralische-fundament-der-usa-beschaedigt-83093

Harsanyi, J. C. (1967). Games with Incomplete Information Played by “Bayesian” Players, I–III, Part I. The Basic Model. *Management Science, 14*(3), 159-182. https://doi.org/10.1287/mnsc.14.3.159

Harsanyi, J. C. (1968a). Games with Incomplete Information Played by “Bayesian” Players, I–III, Part II. Bayesian Equilibrium Points. *Management Science, 14*(5), 320-334. https://doi.org/10.1287/mnsc.14.5.320

Harsanyi, J. C. (1968b). Games with Incomplete Information Played by “Bayesian” Players, I–III, Part III. The Basic Probability Distribution of the Game. *Management Science, 14*(7), 486-502. https://doi.org/10.1287/mnsc.14.7.486

Hartenberger, U., Lorenz, D., & Lützkendorf, T. (2013). A shared built environment professional identity through education and training. *Building Research & Information, 41*(1), 60-76. https://doi.org/10.1080/09613218.2013.736202

Hergueux, J., Jacquemet, N., Luchini, S., & Shogren, J. F. (2019). Leveraging the Honor Code: Public Goods Contributions under Oath. *PSE Working Papers, 2016-22 (revised),* 1-22.

Hoering, S., Kühl, S., & Schulze-Fielitz, A. (2001). Homogenität und Heterogenität in der Gruppenzusammensetzung. Eine mikropolitische Studie über Entscheidungsprozesse in der Gruppenarbeit. *Arbeit, 10*(4), 331-351. https://doi.org/10.1515/arbeit-2001-0406

Hurkens, S., & Kartik, N. (2009). Would I lie to you? On social preferences and lying aversion. *Experimental Economics, 12*(2), 180-192. https://doi.org/10.1007/s10683-008-9208-2

Innocenti, A., & Pazienza, M. G. (2006). Altruism and Gender in the Trust Game. *Labsi Working Papers, 5/2006.*

InvestmentNews. (2019). *SEC charges Massachusetts RIA with fraud*. InvestmentNews. Retrieved 15 October 2020 from https://www.investmentnews.com/sec-charges-massachusetts-ria-with-fraud-80838

Ismayilov, H., & Potters, J. (2016). Why do promises affect trustworthiness, or do they? *Experimental Economics, 19*(2), 382-393. https://doi.org/10.1007/s10683-015-9444-1

Jacquemet, N., Joule, R.-V., Luchini, S., & Shogren, J. F. (2013). Preference elicitation under oath. *Journal of Environmental Economics and Management, 65*(1), 110-132. https://doi.org/10.1016/j.jeem.2012.05.004

Jacquemet, N., Luchini, S., Malézieux, A., & Shogren, J. F. (2020). Who'll stop lying under oath? Empirical evidence from tax evasion games. *European Economic Review, 124*(103369). https://doi.org/10.1016/j.euroecorev.2020.103369

Jacquemet, N., Luchini, S., Rosaz, J., & Shogren, J. F. (2019). Truth Telling Under Oath. *Management Science, 65*(1), 426-438. https://doi.org/10.1287/mnsc.2017.2892

Jacquemet, N., Luchini, S., Shogren, J. F., & Watson, V. (2019). Discrete Choice under Oaths. *Université Paris1 Panthéon-Sorbonne (Post-Print and Working Papers), halshs-02136103, HAL*, 1-20.

Jacquemet, N., Luchini, S., Shogren, J. F., & Zylbersztejn, A. (2018). Coordination with communication under oath. *Experimental Economics, 21*(3), 627-649. https://doi.org/10.1007/s10683-016-9508-x

Jung, J. C., & Park, S. B. (2017). Case Study: Volkswagen's Diesel Emissions Scandal. *Thunderbird International Business Review, 59*(1), 127-137. https://doi.org/10.1002/tie.21876

Kahneman, D. (2003). Maps of Bounded Rationality: Psychology for Behavioral Economics. *American Economic Review, 93*(5), 1449-1475. https://doi.org/10.1257/000282803322655392

Kahneman, D. (2011). *Thinking, fast and slow*. Farrar, Straus and Giroux.

Kaptein, M. (2004). Business Codes of Multinational Firms: What Do They Say? *Journal of Business Ethics, 50*(1), 13-31. https://doi.org/10.1023/B:BUSI.0000021051.53460.da

Kemper, N., Nayga, R. M., Jr., Popp, J., & Bazzani, C. (2016). The Effects of Honesty Oath and Consequentiality in Choice Experiments. *Agricultural and Applied Economics Association, 2016 Annual Meeting, July 31-August 2*. https://doi.org/10.22004/ag.econ.235381

Knack, S., & Keefer, P. (1997). Does Social Capital Have an Economic Payoff? A Cross-Country Investigation. *The Quarterly Journal of Economics, 112*(4), 1251-1288. https://doi.org/10.1162/003355300555475

Knebel, H. (2011). Teamfähigkeit und die Beurteilung von Teamleistungen. In T. R. Hummel & E. Zander (Eds.), *Neuere Entwicklungen in ausgewählten Bereichen der Personalpolitik* (pp. 107-139). Rainer Hampp Verlag.

Kocher, M. G., Schudy, S., & Spantig, L. (2018). I Lie? We Lie! Why? Experimental Evidence on a Dishonesty Shift in Groups. *Management Science, 64*(9), 3995-4008. https://doi.org/10.1287/mnsc.2017.2800

Koessler, A.-K., Torgler, B., Feld, L. P., & Frey, B. S. (2019). Commitment to pay taxes: Results from field and laboratory experiments. *European Economic Review, 115*, 78-98. https://doi.org/10.1016/j.euroecorev.2019.02.006

Kreps, D. M. (1997). Intrinsic Motivation and Extrinsic Incentives. *The American Economic Review, 87*(2), 359-364.

Kretschmer, A. (2002). Maßnahmen zur Kontrolle von Korruption: Eine modelltheoretische Untersuchung. *Arbeitspapiere des Instituts für Genossenschaftswesen der Westfälischen Wilhelms-Universität Münster, 25*.

Kroher, M., & Wolbring, T. (2015). Social control, social learning, and cheating: Evidence from lab and online experiments on dishonesty. *Social Science Research, 53*, 311-324. https://doi.org/10.1016/j.ssresearch.2015.06.003

Kugler, T., Bornstein, G., Kocher, M. G., & Sutter, M. (2007). Trust between individuals and groups: Groups are less trusting than individuals but just as trustworthy. *Journal of Economic Psychology, 28*(6), 646-657. https://doi.org/10.1016/j.joep.2006.12.003

Kuroki, M. (2011). Does Social Trust Increase Individual Happiness in Japan? *The Japanese Economic Review, 62*(4), 444-459. https://doi.org/10.1111/j.1468-5876.2011.00533.x

Lang, M., Mitkidis, P., Kundt, R., Nichols, A., Krajčíková, L., & Xygalatas, D. (2016). Music As a Sacred Cue? Effects of Religious Music on Moral Behavior. *Frontiers in Psychology, 7*(814), 1-13. https://doi.org/10.3389/fpsyg.2016.00814

Leland, H. E., & Pyle, D. H. (1977). Informational Asymmetries, Financial Structure, and Financial Intermediation. *The Journal of Finance, 32*(2), 371-387. https://doi.org/10.2307/2326770

Li, S., Bühren, C., Frank, B., & Qin, H. (2015). Group Decision Making in a Corruption Experiment: China and Germany Compared. *Journal of Economics and Statistics, 235*(2), 207-227. https://doi.org/10.1515/jbnst-2015-0207

List, J. A., & Momeni, F. (2020). When Corporate Social Responsibility Backfires: Evidence from a Natural Field Experiment. *Management Science*, Forthcoming. https://doi.org/10.1287/mnsc.2019.3540

Loonen, T., & Rutgers, M. R. (2017). Swearing to be a good banker: Perceptions of the obligatory banker's oath in the Netherlands. *Journal of Banking Regulation, 18*(1), 28-47. https://doi.org/10.1057/jbr.2016.2

López-Pérez, R., & Spiegelman, E. (2013). Why do people tell the truth? Experimental evidence for pure lie aversion. *Experimental Economics, 16*(3), 233-247. https://doi.org/10.1007/s10683-012-9324-x

Lundquist, T., Ellingsen, T., Gribbe, E., & Johannesson, M. (2009). The aversion to lying. *Journal of Economic Behavior & Organization, 70*(1-2), 81-92. https://doi.org/10.1016/j.jebo.2009.02.010

Lynch, S. N. (2016). *Wells Fargo Under Criminal Investigation in California Over Phony Accounts Scandal*. Time Business. Retrieved 15 October 2020 from http://time.com/4537718/wells-fargo-under-criminal-investigation/

Mahon, J. E. (2008). Two Definitions of Lying. *International Journal of Applied Philosophy, 22*(2), 211-230. https://doi.org/10.5840/ijap200822216

Majolo, B., Ames, K., Brumpton, R., Garratt, R., Hall, K., & Wilson, N. (2006). Human friendship favours cooperation in the Iterated Prisoner's Dilemma. *Behaviour, 143*(11), 1383-1395. https://doi.org/10.1163/156853906778987506

Mazar, N., Amir, O., & Ariely, D. (2008). The Dishonesty of Honest People: A Theory of Self-Concept Maintenance. *Journal of Marketing Research, 45*(6), 633-644. https://doi.org/10.1509/jmkr.45.6.633

McCord, L. B., Greenhalgh, K., & Magasin, M. (2004). Businesspersons Beware: Lying is a Crime. *Graziadio Business Review, 7*(3). https://gbr.pepperdine.edu/2010/08/businesspersons-beware-lying-is-a-crime/

McKnight, H. D., & Chervany, N. L. (2001). Trust and Distrust Definitions: One Bite at a Time. In R. Falcone, M. Singh, & Y.-H. Tan (Eds.), *Trust in Cyber-societies: Integrating the Human and Artificial Perspectives* (Vol. 1, pp. 27-54). Springer. https://doi.org/10.1007/3-540-45547-7_3

Morgan, P. M., & Tindale, R. S. (2002). Group vs Individual Performance in Mixed-Motive Situations: Exploring an Inconsistency. *Organizational Behavior and Human Decision Processes, 87*(1), 44-65. https://doi.org/10.1006/obhd.2001.2952

Muehlheusser, G., Roider, A., & Wallmeier, N. (2015). Gender differences in honesty: Groups versus individuals. *Economics Letters, 128*, 25-29. https://doi.org/10.1016/j.econlet.2014.12.019

O'Leary, C., & Pangemanan, G. (2007). The Effect of Groupwork on Ethical Decision-Making of Accountancy Students. *Journal of Business Ethics, 75*(3), 215-228. https://doi.org/10.1007/s10551-006-9248-5

OECD. (2007). *Integrity in Public Procurement: Good Practice from A to Z*. OECD Publishing. https://doi.org/10.1787/9789264027510-en

Pe'er, E., & Feldman, Y. (2020). Honesty Pledges for the Behaviorally-based Regulation of Dishonesty. *Working Paper*. https://doi.org/10.2139/ssrn.3615743

Peeters, R., Vorsatz, M., & Walzl, M. (2013). Truth, Trust, and Sanctions: On Institutional Selection in Sender-Receiver Games. *The Scandinavian Journal of Economics, 115*(2), 508-548. https://doi.org/10.1111/sjoe.12003

Peeters, R., Vorsatz, M., & Walzl, M. (2015). Beliefs and truth-telling: A laboratory experiment. *Journal of Economic Behavior & Organization, 113*, 1-12. https://doi.org/10.1016/j.jebo.2015.02.009

Pruckner, G. J., & Sausgruber, R. (2013). Honesty on the Streets: A Field Study on Newspaper Purchasing. *Journal of the European Economic Association, 11*(3), 661-679. https://doi.org/10.1111/jeea.12016

Robbins, S. P., & Judge, T. A. (2013). *Organizational behavior* (15 ed.). Pearson Education.

Rose-Ackerman, S. (1975). The economics of corruption. *Journal of Public Economics, 4*(2), 187-203. https://doi.org/10.1016/0047-2727(75)90017-1

Rosenthal, R. W. (1981). Games of perfect information, predatory pricing and the chain-store paradox. *Journal of Economic Theory, 25*(1), 92-100. https://doi.org/10.1016/0022-0531(81)90018-1

Rothenhäusler, D., Schweizer, N., & Szech, N. (2018). Guilt in voting and public good games. *European Economic Review, 101*, 664-681. https://doi.org/10.1016/j.euroecorev.2017.08.001

Rotter, J. B. (1967). A new scale for the measurement of interpersonal trust. *Journal of Personality, 35*(4), 651-665. https://doi.org/10.1111/j.1467-6494.1967.tb01454.x

Rousselet, G. A., Pernet, C. R., & Wilcox, R. R. (2019). A practical introduction to the bootstrap: a versatile method to make inferences by using data-driven simulations. *Working Paper*. https://doi.org/10.31234/osf.io/h8ft7

Rubinstein, A. (2007). Instinctive and Cognitive Reasoning: A Study of Response Times. *The Economic Journal*, *117*(523), 1243-1259. https://doi.org/10.1111/j.1468-0297.2007.02081.x

Rutgers, M. R. (2010). The Oath of Office as Public Value Guardian. *The American Review of Public Administration*, *40*(4), 428-444. https://doi.org/10.1177/0275074009349598

Rutgers, M. R. (2013). Will the Phoenix Fly Again? *Review of Social Economy, 71*(2), 249-276. https://doi.org/10.1080/00346764.2013.800306

Ryan, R. M., & Deci, E. L. (2000). Intrinsic and Extrinsic Motivations: Classic Definitions and New Directions. *Contemporary Educational Psychology, 25*(1), 54-67. https://doi.org/10.1006/ceps.1999.1020

Sánchez-Pagés, S., & Vorsatz, M. (2007). An experimental study of truth-telling in a sender-receiver game. *Games and Economic Behavior*, *61*(1), 86-112. https://doi.org/10.1016/j.geb.2006.10.014

Sapienza, P., Toldra-Simats, A., & Zingales, L. (2013). Understanding Trust. *The Economic Journal, 123*(573), 1313-1332. https://doi.org/10.1111/ecoj.12036

Schickora, J. T. (2011). Bringing the Four-Eyes-Principle to the Lab. *Munich Discussion Paper 2011, 3*. https://doi.org/10.5282/ubm/epub.12160

Schulze, G. G., & Frank, B. (2003). Deterrence versus intrinsic motivation: Experimental evidence on the determinants of corruptibility. *Economics of Governance, 4*(2), 143-160. https://doi.org/10.1007/s101010200059

Schumacher, C. (2006). Trust--A Source of Success in Strategic Alliances? *Schmalenbach Business Review, 58*(3), 259-278. https://doi.org/10.1007/bf03396734

Schunk, D., & Betsch, C. (2006). Explaining heterogeneity in utility functions by individual differences in decision modes. *Journal of Economic Psychology, 27*(3), 386-401. https://doi.org/10.1016/j.joep.2005.08.003

Securities and Exchange Commission. (2001). *Litigation Release No. 16925 / March 8, 2001: SEC v. Yun Soo Oh Park and Tokyo Joe's Societe Anonyme Corp., N.D. Ill., Civil No. 00-C-0049, filed January 5, 2000*. Securities and Exchange Commission. Retrieved 15 October 2020 from https://www.sec.gov/litigation/litreleases/lr16925.htm

Securities and Exchange Commission. (2008). *SEC Charges Bernard L. Madoff for Multi-Billion Dollar Ponzi Scheme*. Securities and Exchange Commission. Retrieved 15 October 2020 from https://www.sec.gov/news/press/2008/2008-293.htm

Securities and Exchange Commission. (2019). *SEC Charges Investment Adviser With Fraud*. Securities and Exchange Commission. Retrieved 15 October 2020 from https://www.sec.gov/news/press-release/2019-77

Servátka, M., Tucker, S., & Vadovič, R. (2011). Building Trust--One Gift at a Time. *Games, 2*(4), 412-433. https://doi.org/10.3390/g2040412

Shalvi, S., Gino, F., Barkan, R., & Ayal, S. (2015). Self-Serving Justifications: Doing Wrong and Feeling Moral. *Current Directions in Psychological Science, 24*(2), 125-130. https://doi.org/10.1177/0963721414553264

Sobel, J. (2009). Signaling Games. In R. A. Meyers (Ed.), *Encyclopedia of Complexity and Systems Science* (pp. 8125-8139). Springer.

Soraperra, I., Weisel, O., Zultan, R. i., Kochavi, S., Leib, M., Shalev, H., & Shalvi, S. (2017). The bad consequences of teamwork. *Economics Letters, 160*, 12-15. https://doi.org/10.1016/j.econlet.2017.08.011

Sulmasy, D. P. (1999). What is an oath and why should a physician swear one? *Theoretical Medicine and Bioethics, 20*(4), 329-346. https://doi.org/10.1023/a:1009968512510

Sutter, M. (2009). Deception Through Telling the Truth?! Experimental Evidence from Individuals and Teams. *The Economic Journal, 119*(534), 47-60. https://doi.org/10.1111/j.1468-0297.2008.02205.x

Teixeira, A. A. C., & Rocha, M. F. (2010). Cheating by economics and business undergraduate students: an exploratory international assessment. *Higher Education, 59*(6), 663-701. https://doi.org/10.1007/s10734-009-9274-1

UNIDO. (2020). *What is the four-eyes principle?* UNIDO. Retrieved 15 October 2020 from https://www.unido.org/overview/member-states/change-management/faq/what-four-eyes-principle

Utikal, V., & Fischbacher, U. (2013). Disadvantageous lies in individual decisions. *Journal of Economic Behavior & Organization, 85*, 108-111. https://doi.org/10.1016/j.jebo.2012.11.011

van der Linden, B. (2013). Principles as “Rules Of Thumb”: A Particularist Approach to Codes of Ethics and an Analysis of the Dutch Banking Code. *Review of Social Economy, 71*(2), 209-227. https://doi.org/10.1080/00346764.2013.799966

Vanberg, C. (2008). Why Do People Keep Their Promises? An Experimental Test of Two Explanations. *Econometrica, 76*(6), 1467-1480. https://doi.org/10.3982/ecta7673

Vanberg, C. (2017). Who never tells a lie? *Experimental Economics, 20*(2), 448-459. https://doi.org/10.1007/s10683-016-9491-2

Vranceanu, R., & Dubart, D. (2019). Deceitful communication in a sender-receiver experiment: Does everyone have a price? *Journal of Behavioral and Experimental Economics, 79*, 43-52. https://doi.org/10.1016/j.socec.2019.01.005

Waddock, S. A., Bodwell, C., & Graves, S. B. (2002). Responsibility: The new business imperative. *Academy of Management Perspectives*, *16*(2), 132-148. https://doi.org/10.5465/ame.2002.7173581

Warner, S. L. (1965). Randomized Response: A Survey Technique for Eliminating Evasive Answer Bias. *Journal of the American Statistical Association, 60*(309), 63-69. https://doi.org/10.2307/2283137

Wegge, J. (2004). *Führung von Arbeitsgruppen* (1 ed.). Hogrefe Verlag.

Weisel, O., & Shalvi, S. (2015). The collaborative roots of corruption. *Proceedings of the National Academy of Sciences*, *112*(34), 10651-10656. https://doi.org/10.1073/pnas.1423035112

Wilcox, R. R. (2016). Comparing dependent robust correlations. *British Journal of Mathematical and Statistical Psychology, 69*(3), 215-224. https://doi.org/10.1111/bmsp.12069

Wouda, J., Bijlstra, G., Frankenhuis, W. E., & Wigboldus, D. H. J. (2017). The Collaborative Roots of Corruption? A Replication of Weisel & Shalvi (2015). *Collabra: Psychology, 3*(1), 1-3. https://doi.org/10.1525/collabra.97

Wu, D., Loke, I. C., Xu, F., & Lee, K. (2011). Neural correlates of evaluations of lying and truth-telling in different social contexts. *Brain Research, 1389*, 115-124. https://doi.org/10.1016/j.brainres.2011.02.084

Dissertation declaration

"I herewith give assurance that I completed this dissertation independently without prohibited assistance of third parties or aids other than those identified in this dissertation. All passages that are drawn from published or unpublished writings, either word-for-word or in paraphrase, have been clearly identified as such. Third parties were not involved in the drafting of the content of this dissertation; most specifically, I did not employ the assistance of a dissertation advisor. No part of this thesis has been used in another doctoral or tenure process."