

Beck, Tobias

Conference Paper

Size matters! Lying and Mistrust in the Continuous Deception Game

Beiträge zur Jahrestagung des Vereins für Socialpolitik 2020: Gender Economics

Provided in Cooperation with:

Verein für Socialpolitik / German Economic Association

Suggested Citation: Beck, Tobias (2020) : Size matters! Lying and Mistrust in the Continuous Deception Game, Beiträge zur Jahrestagung des Vereins für Socialpolitik 2020: Gender Economics, ZBW - Leibniz Information Centre for Economics, Kiel, Hamburg

This Version is available at:

<https://hdl.handle.net/10419/224530>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

Size matters!

Lying and Mistrust in the Continuous Deception Game

Tobias Beck

Institute of Economics, University of Kassel, Nora Platiel-Straße 4, 34127 Kassel, Hessen, Germany;
tobi.beck@gmx.net; ORCID: 0000-0002-4162-3648

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.

Conflicts of interest: The author declares no conflict of interest.

Funding: This research received no external funding.

Availability of data and material: The datasets generated and analyzed during the current study are available from the author on reasonable request.

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 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 2400 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.¹

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.

¹ 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 for 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*.² 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 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

² 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 the one 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 the one of the receiver).³

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.⁴ 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.⁵ 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

³ A similar distinction is also used by Gneezy (2005).

⁴ 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.

⁵ 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.⁶ The resulting binary payoff structure of the sender greatly limits the possibilities of comparing two lies with different sizes to one another, if both of them 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.⁷

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.⁸ 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 (as

⁶ Lundquist et al. (2009) determined the talent score of the sellers based on a test that took place before the actual experiment started.

⁷ 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.

⁸ 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.

defined by Sobel, 2009) that is framed as an investment game.⁹ 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.¹⁰ 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 :

$$\pi_A(i) = m_A * i \quad (1)$$

with $m_A > 0$.¹¹

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

⁹ 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.

¹⁰ The optimal investment i^* is randomly determined by a uniform distribution: $P([0, i^*]) = \frac{i^*}{i_{max}}$ with $i^* \in [0, i_{max}]$.

¹¹ The factor m_A is the advisor's payoff factor. This factor determines to which extent he or she profits from the investment i .

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

with $m_I > 0$.¹²

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 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.

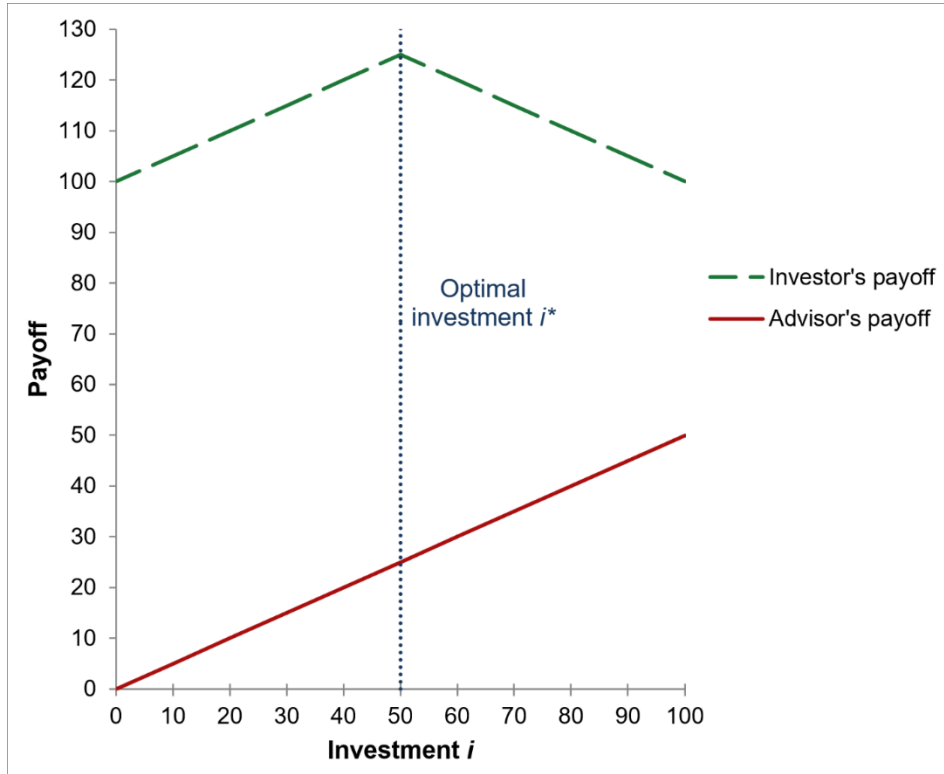


Figure 1. Payoff structure of the CDG

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 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.

¹² 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^* .

3.2. The game process

Figure 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, the advisor is informed about the value of the optimal investment i^* . Afterwards, he or she is instructed to report this optimum to their 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 .¹³ 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.

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^* .¹⁴ This guess i_{guess}^* reflects the investor's first-order beliefs about the advisor. Note that 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.

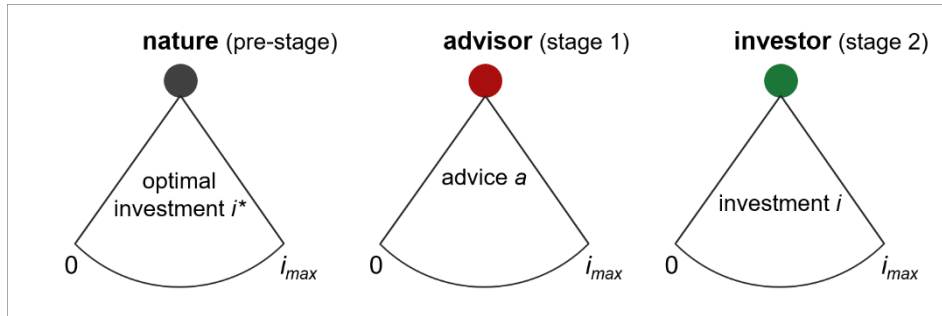


Figure 2. Game tree of the CDG

¹³ 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).

¹⁴ 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).

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.

Note that, 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).¹⁵

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:

$$L = \frac{a - i^*}{i^*} \quad (3)$$

with $i^* > 0$.¹⁶

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 $a > i^*$: $L > 0$).

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

$$L_{guess} = \frac{a - i_{guess}^*}{i_{guess}^*} \quad (4)$$

with $i_{guess}^* > 0$.

¹⁵ 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).

¹⁶ 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.

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:

$$\bar{T} = \frac{i - a}{a} \quad (5)$$

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.¹⁷

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:

$$\bar{T}_{guess} = \frac{i_{guess} - a}{a} \quad (6)$$

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

¹⁷ 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.

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:

$$F_{A;guess} = \frac{i_{guess} - i^*}{i^*} \quad (7)$$

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:

$$F_{I;guess} = \frac{i - i_{guess}^*}{i_{guess}^*} \quad (8)$$

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:

$$F = \frac{i - i^*}{i^*} \quad (9)$$

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 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{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,¹⁸ 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 $F_{A;guess}$).

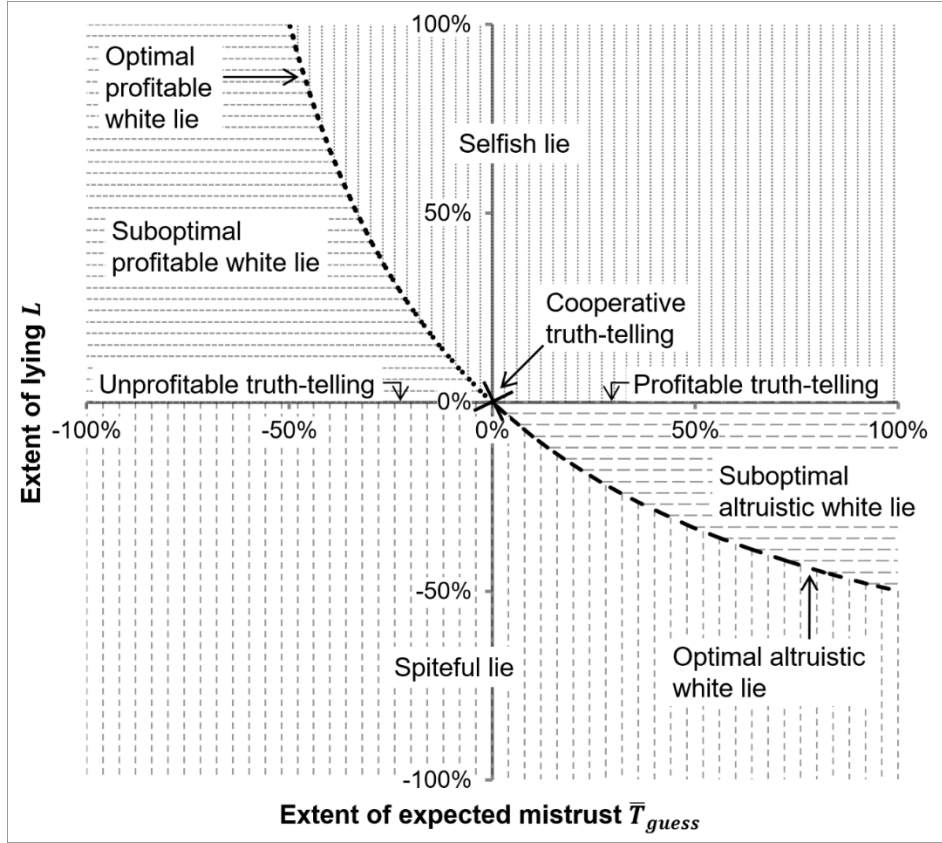


Figure 3. Taxonomy of lies and truth-telling for the advisor

As summarized in Table 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 ($F_{A;guess} = 0$).

¹⁸ As described above, Erat and Gneezy (2012) differentiate between four types of lies: altruistic white lies (i.e., lies that are expected to reduce the sender's payoff while increasing the one 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 the one of the receiver).

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 the one of the investor.

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

Table 1. Definition of classes of advisor strategies

3.4.2. Taxonomy of mistrust and trust

Figure 4 displays the taxonomy of mistrust and trust for the investor based on their percentage extent of mistrust (\bar{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.¹⁹

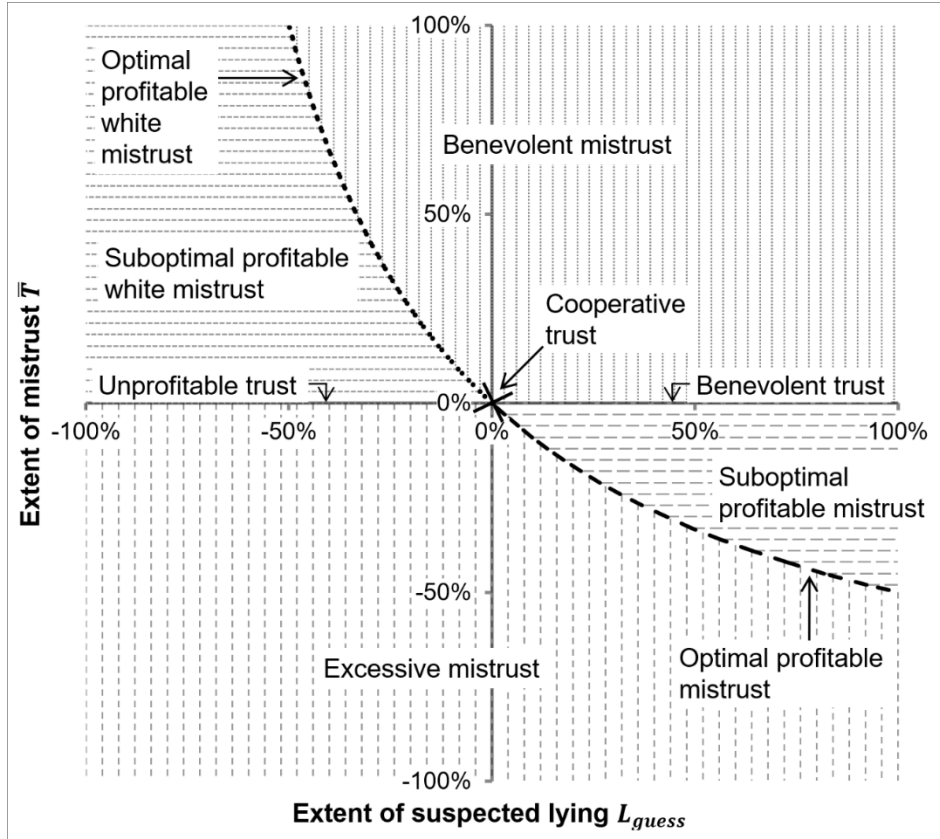


Figure 4. Taxonomy of mistrust and trust for the investor

As summarized in Table 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 ($F_{I;guess} = 0$).

¹⁹ 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 the one 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 ($F_{I,guess} < 0$)	<i>Excessive mistrust</i>	<i>Unprofitable trust</i>	<i>Suboptimal profitable white mistrust</i>
	Optimal investment ($F_{I,guess} = 0$)	<i>Optimal profitable mistrust</i>	<i>Cooperative trust</i>	<i>Optimal profitable white mistrust</i>
	Overinvestment ($F_{I,guess} > 0$)	<i>Suboptimal profitable mistrust</i>	<i>Benevolent trust</i>	<i>Benevolent mistrust</i>

Table 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 A 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.²⁰ 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 A shows that, on the one hand, *rational advisors* would engage in either selfish lying, optimal profitable white lying, or cooperative truth-telling (with $F_{A;guess} \geq 0$ and $\bar{T}_{guess} \leq 0$).²¹ 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 $L \geq 0$).²² 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 $\bar{T} \leq 0$).²³ 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 $\bar{T} \leq 0$).²⁴

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.

²⁰ As discussed in the literature section, 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 A.

²¹ For proof, see appendices A.2.1. to A.2.3.

²² For proof, see appendix A.2.4.

²³ For proof, see appendices A.2.1. to A.2.3.

²⁴ For proof, see appendix A.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. It yields 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 0.5.²⁵ 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.²⁶ 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.²⁷ 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,

²⁵ 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.

²⁶ To simplify the amounts of the optimal investments, I only allowed optimal investments that were divisible by 5. 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, I find in appendix D 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.

²⁷ 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' average extent of lying and the investors' average extent of suspected lying at the 5%-threshold. 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*. 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. 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.²⁸ 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' 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

²⁸ If the participants did not complete their questionnaire, they received only half of their original payoff.

real extent of misinvestment (7.1.3.). Note that all values in these subsections refer to averages over all ten rounds.

7.1.1. Extent of (suspected) lying

Table 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.²⁹ Secondly, it displays the average percentage extent of (suspected) lying over all ten rounds.

Concerned variables	Lying (advisor)	Suspected lying (investor)	Difference of averages (1 st -2 nd)	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. 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.³⁰ To show why, Figure 5 visualizes the distributions of the percentage extent of lying (L) on the left (5a) and of suspected lying (L_{guess}) on the right (5b).

²⁹ 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.

³⁰ 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.

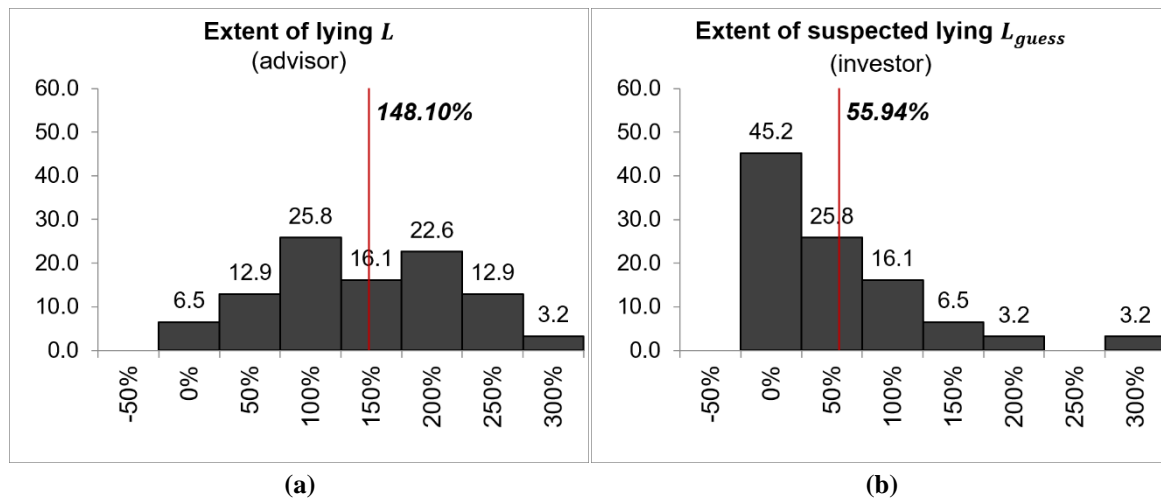


Figure 5. Distributions of the extent of lying and the extent of suspected lying: **(a)** Extent of lying L ; **(b)** Extent of suspected lying L_{guess}

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 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.³¹ 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 (2 nd -1 st)	p-value (two-sided Mann-Whitney U test ¹)
Proportion of (expected) mistrust	68.71%	72.26%	3.55%	0.701
<i>Proportion of (expected) risk-reducing mistrust</i>	58.06%	57.10%	-0.96%	1.000
<i>Proportion of (expected) risk-seeking mistrust</i>	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 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 6 illustrates the distributions of the advisors' percentage extent of expected mistrust (\bar{T}_{guess}) on the left (6a) and the investors' percentage extent of mistrust (\bar{T}) on the right (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 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-

³¹ 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 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.³² 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 (1 st -2 nd)	p-value (two-sided Mann-Whitney U test ¹)
Proportion of expected misinvestments	73.22%	57.42%	15.80%	0.086
<i>Proportion of expected underinvestments</i>	11.61%	28.39%	-16.78%	0.008
<i>Proportion of expected overinvestments</i>	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 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 underinvestments (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; Mann-Whitney U test: $p < 0.001$). Going into more detail, Figure 7 visualizes the distributions of the percentage extent of expected misinvestment of the advisors ($F_{A,guess}$) on the left (7a) and of the investors ($F_{I,guess}$) on the right (7b).

³² 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.

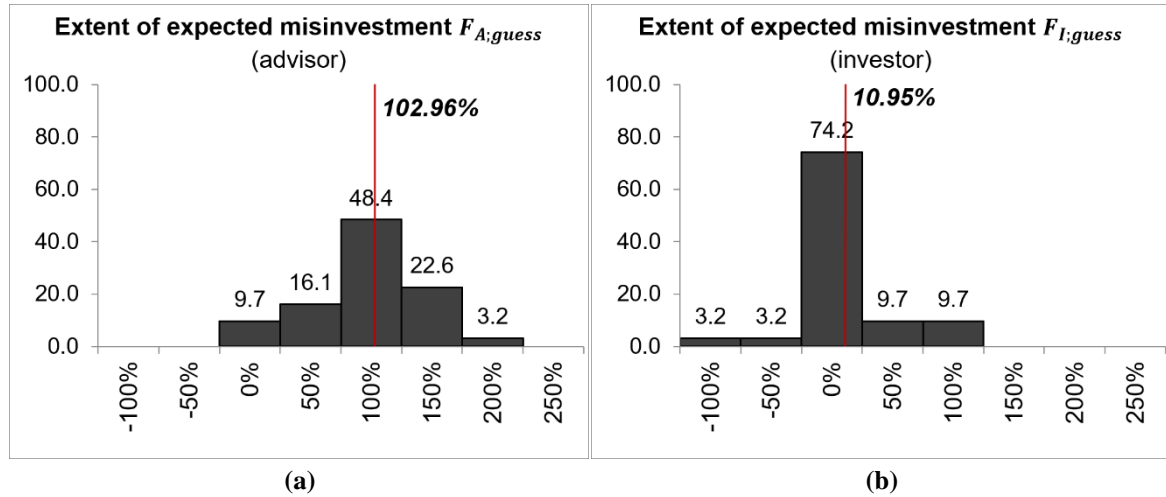


Figure 7. Distributions of both players' extent of expected misinvestment: **(a)** Extent of expected misinvestment of the advisor $F_{A;guess}$; **(b)** Extent of expected misinvestment of the investor $F_{I;guess}$

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 6. Proportion and extent of misinvestment

In order to assess the quality of both players' estimates of the outcomes of the investments, Table 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.³³ 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.

³³ 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.

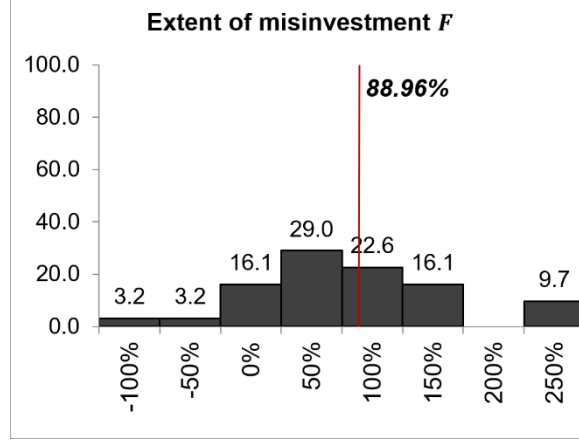


Figure 8. Distribution of the extent of misinvestment F

For a more detailed analysis, Figure 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.).

7.2.1. Lying and expected mistrust (advisors)

Figure 9 visualizes the relationship between the advisors' lying behavior and their first-order beliefs about their investors.³⁴ 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

³⁴ For the purpose of illustration, only the most relevant area of the plot in Figure 9 is displayed. Also note that the dotted line in Figure 9 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 underinvestments* ($F_{A,guess} < 0$), points above it represent *expectations of overinvestments* ($F_{A,guess} > 0$).

between L and $-\bar{T}_{guess}$: $\rho = 0.397$ 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; Fisher exact test: $p < 0.001$). 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; Fisher exact 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}).*

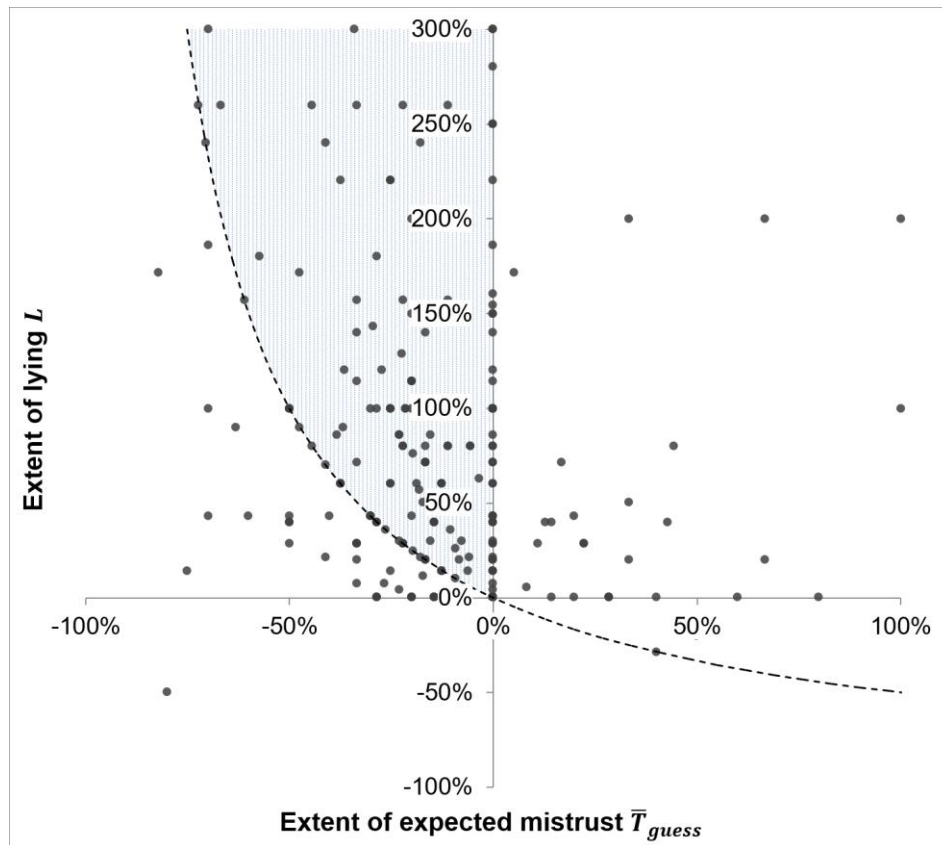


Figure 9. Relationship between the extent of lying L and the extent of expected mistrust \bar{T}_{guess}

To show which strategies the advisors pursued, Table 7 summarizes the proportions of *classes of observed advisor strategies* (which are also displayed in a non-aggregated form in Figure 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 ($L > 0$)
Underinvestment ($F_{A,guess} < 0$)	<i>Spiteful lie</i> 0.32%	<i>Unprofitable truth-telling</i> 4.19%	<i>Suboptimal profitable white lie</i> 7.10%
Optimal investment ($F_{A,guess} = 0$)	<i>Optimal altruistic white lie</i> 0.65%	<i>Cooperative truth-telling</i> 13.87%	<i>Optimal profitable white lie</i> 12.26%
Overinvestment ($F_{A,guess} > 0$)	<i>Suboptimal altruistic white lie</i> 0.32%	<i>Profitable truth-telling</i> 3.55%	<i>Selfish lie</i> 57.74%
Total	Understating lie 1.29%	Truth-telling 21.61%	Overstating lie 77.10%

Table 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 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.³⁵ It comes as no surprise that the advisors pursued these

³⁵ 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.

rational strategies more often than other strategies in every round (two-sided Binomial tests: $p < 0.001$ for each round³⁶). 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.*

7.2.2. Mistrust and suspected lying (investors)

The scatter plot in Figure 10 visualizes the relationship between the investors' mistrusting behavior and their first-order beliefs about their advisors.³⁷ 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; Fisher exact 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; Fisher exact 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}).*

³⁶ 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: $p_{A;rat}(i^*) = \frac{1}{i_{max}(i_{max}-i^*)} * [0.5 * i_{max}^2 + 0.5 * i^{*2} - i_{max} * i^*]$.

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.

³⁷ For the purpose of illustration, only the most relevant area of the plot in Figure 10 is displayed. Also note that the dotted line in Figure 10 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 underinvestments* ($F_{I;guess} < 0$), whereas points above it represent *expectations of overinvestments* ($F_{I;guess} > 0$).

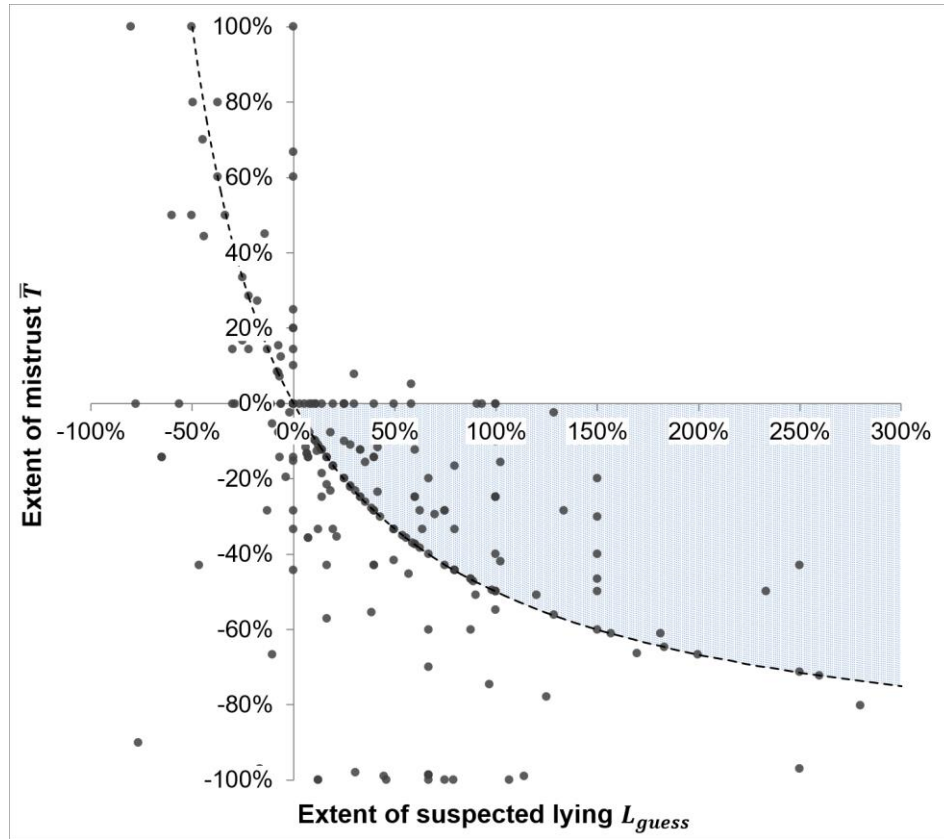


Figure 10. Relationship between the extent of mistrust \bar{T} and the extent of suspected lying L_{guess}

To illustrate which strategies the investors chose, Table 8 provides an overview of the proportions of *classes of observed investor strategies* (which are also displayed in a non-aggregated form in Figure 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 ($F_{I,guess} < 0$)	<i>Excessive mistrust</i> 20.65%	<i>Unprofitable trust</i> 4.19%
			<i>Suboptimal profitable white mistrust</i> 3.55%
	Optimal investment ($F_{I,guess} = 0$)	<i>Optimal profitable mistrust</i> 20.65%	<i>Cooperative trust</i> 17.10%
			<i>Optimal profitable white mistrust</i> 4.84%
	Overinvestment ($F_{I,guess} > 0$)	<i>Suboptimal profitable mistrust</i> 15.80%	<i>Benevolent trust</i> 6.45%
			<i>Benevolent mistrust</i> 6.77%
	Total	Risk-reducing mistrust 57.10%	Trusting behavior 27.74%
			Risk-seeking mistrust 15.16%

Table 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.³⁸ 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 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 Binomial test³⁹: $p < 0.001$). This finding is consistent with *hypothesis H2*, which speaks in favor of the conjecture that the investors engaged in rational decision making based on individual trust preferences and rational beliefs.

³⁸ 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 10, this refers to all strategy points that are located on the dotted line except for the ones at the point of origin.

³⁹ 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 $a = i_{max}$. With $i_{max} = 100$, it yields: $p_{I,rat}(a) \leq p_{I,rat}(a = i_{max} = 100) = 0.5$. Since not all investors received the same advice, I used this upper limit of $p_{I,rat}$ to perform the most conservative two-sided Binomial test possible that compares the proportion of rational strategies to its expected value based on random choices.

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.*

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 their potential to manipulate their investors into account (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 $p = 0.828$). The difference between these two correlations is significant (two-sided Fisher's Z test: $p < 0.001$). 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.*

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 $p = 0.716$). The difference between these two correlations is significant (two-sided Fisher's Z 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.⁴⁰ On this basis, in appendix C, 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 C2 in appendix C*). 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

⁴⁰ 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 C.

higher preference for risk (*see Finding C1*). 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 C4*). 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 C3*). 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.⁴¹ 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). 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.).

⁴¹ Both proportions were approximately 78%.

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.⁴² 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 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 disproportionately 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

⁴² It should be reminded that Gneezy (2005) and I use the same ratio between the profits from lying to the senders 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 the one of Gneezy (2005).

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).⁴³

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 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-

⁴³ 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 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.

sized lies in particular and when analyzing the intent behind lying or any other deceptive behavior in general.⁴⁴

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 would yield a Pareto improvement. This in turn is in support of Vanberg (2017), as this means that

⁴⁴ 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.

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 for 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 for themselves and for another person might not really care about the other person but use the fact that they are helping that person to rationalize the lie.

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 allow drawing conclusions on why people trust or mistrust potential liars.⁴⁵

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*.

⁴⁵ 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.⁴⁶ 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, they 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 a trade-off between monetary gain (or altruism) and risk aversion. If this was the case, however, it would

⁴⁶ 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.

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.

Acknowledgements

I would like to thank Björn Frank for his valuable feedback on the paper.

References

- 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>
- 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>
- 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>
- 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., 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 8 April 2020 from <https://www.belfasttelegraph.co.uk/news/uk/bank-fined-105m-over-mis-selling-28688514.html>
- 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>

- 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>
- 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>
- 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](https://doi.org/10.1016/s0899-8256(03)00119-2)
- 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>
- 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>
- 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).
- 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>
- 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>
- 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., & Paziienza, 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 8 April 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., 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., & Zylbersztejn, A. (2018). Coordination with communication under oath. *Experimental Economics*, 21(3), 627-649. <https://doi.org/10.1007/s10683-016-9508-x>
- 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>
- 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>

- 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>
- 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>
- 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>
- 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
- 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>
- 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>
- 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/eco.j.12036>
- 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 25 August 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 8 April 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 8 April 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>
- Sobel, J. (2009). Signaling Games. In R. A. Meyers (Ed.), *Encyclopedia of Complexity and Systems Science* (pp. 8125-8139). Springer.
- 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>
- 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>
- 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>

Appendices

This section consists of four separate appendices. Firstly, appendix A shows the derivation of the set of game theoretical equilibria of the Continuous Deception Game (CDG). Secondly, appendix B contains the instructions and input screens that I presented to the subjects in the CDG. Thirdly, appendix C 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. Finally, appendix D analyzes the temporal consistency of both players' behavior and first-order beliefs.

Appendix A. 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.⁴⁷ 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 (A.1.). In a second step, I derive the set of game theoretical equilibria with monetary and non-monetary motivation (A.2.).

A.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 $i \in [0, i_{max}]$:

$$\pi_I(i) = \begin{cases} i \leq i^*: & i_{max} + m_I * i \\ i > i^*: & i_{max} + m_I * (2 * i^* - i) \end{cases} \quad (10)$$

with $m_I > 0$.

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

⁴⁷ 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).

investment i^* is located somewhere between 0 and the maximal investment i_{max} with equal probability. On this basis, the *investor's expected payoff* $\pi_{I;hoec}^e(i)$ can be defined as the following function of the investment i :

$$\pi_{I;hoec}^e(i) = \frac{1}{i_{max}} \int_0^{i_{max}} \pi_I(i) di^* \quad (11)$$

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 ($i^* \geq i$). It yields:

$$\begin{aligned} \pi_{I;hoec}^e(i) &= \frac{1}{i_{max}} \int_0^{i_{max}} \pi_I(i) di^* \\ &= \frac{1}{i_{max}} * \lim_{b \nearrow i} \left(\int_0^b \underbrace{\pi_I(b)}_{i^* < i} di^* \right) + \frac{1}{i_{max}} \int_i^{i_{max}} \underbrace{\pi_I(i)}_{i^* \geq i} di^* \\ &= \frac{1}{i_{max}} * \lim_{b \nearrow i} \left(\int_0^b (i_{max} + m_I * (2 * i^* - b)) di^* \right) + \frac{1}{i_{max}} \int_i^{i_{max}} (i_{max} + m_I * i) di^* \\ &= \frac{1}{i_{max}} \int_0^i (i_{max} + m_I * (2 * i^* - i)) di^* + \frac{1}{i_{max}} \int_i^{i_{max}} (i_{max} + m_I * i) di^* \quad (12) \\ &= \frac{1}{i_{max}} * [i_{max} * i^* + m_I * (i^{*2} - i * i^*)]_0^i + \frac{1}{i_{max}} * [(i_{max} + m_I * i) * i^*]_i^{i_{max}} \\ &= \frac{1}{i_{max}} * [i_{max} * i + m_I * (i^2 - i^2)] + \frac{1}{i_{max}} * (i_{max} - i) * (i_{max} + m_I * i) \\ &= \frac{i}{i_{max}} * i_{max} + \left(1 - \frac{i}{i_{max}}\right) * (i_{max} + m_I * i) \\ &= i_{max} + m_I * \left(i - \frac{i^2}{i_{max}}\right) \end{aligned}$$

with $m_I > 0$.

Given this, the investor seeks to maximize their expected payoff $\pi_{I;hoec}^e(i)$ 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 $\pi_{I;hoec}^e(i)$ with respect to the investment i , leads to:

$$\begin{aligned}
\max_i(\pi_{I;hoec}^e(i)) &\rightarrow \frac{\partial \pi_{I;hoec}^e(i)}{\partial i} \stackrel{!}{=} 0 \\
&\leftrightarrow \frac{\partial \left(i_{max} + m_I * \left(i - \frac{i^2}{i_{max}} \right) \right)}{\partial i} = 0 \\
&\leftrightarrow m_I - \frac{2 * m_I * i}{i_{max}} = 0 \\
&\leftrightarrow i_{BR;hoec} = \frac{i_{max}}{2}
\end{aligned} \tag{13}$$

with $m_I > 0$.

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

$$\begin{aligned}
\pi_{I;BR;hoec}^e &= \pi_{I;hoec}^e(i = i_{BR;hoec}) \\
&= i_{max} + m_I * \left(i_{BR;hoec} - \frac{i_{BR;hoec}^2}{i_{max}} \right) \\
&= i_{max} + m_I * \left(\frac{i_{max}}{2} - \left(\frac{i_{max}}{2} \right)^2 * \frac{1}{i_{max}} \right) \\
&= i_{max} * \left(1 + \frac{m_I}{4} \right)
\end{aligned} \tag{14}$$

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. It yields:

$$S_{A;BR;hoec} = \{a | a \in [0, i_{max}]\}. \tag{15}$$

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

$$\pi_{A;BR;hoec}^e = \pi_A(i = i_{BR;hoec}) = m_A * i_{BR;hoec} = i_{max} * \frac{m_A}{2} \quad (16)$$

with $m_A > 0$.

Since all advice a will result in this same expected payoff $\pi_{A;BR;hoec}^e$, this makes every advice $a \in [0, i_{max}]$ 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:

$$S_{equ;hoec} = \left\{ (a_{hoec}^s, i_{hoec}^s) \mid a_{hoec}^s \in [0, i_{max}] \wedge i_{hoec}^s = \frac{i_{max}}{2} \right\}. \quad (17)$$

A.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^* .⁴⁸ 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 :

$$U_A(a, i) = \pi_A(i) - C_L(a). \quad (18)$$

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.

⁴⁸ 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 $C_{\bar{T}}(a, i) \dots$

...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 :

$$U_I(a, i) = \pi_I(i) - C_{\bar{T}}(a, i). \quad (19)$$

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.⁴⁹

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 $(P([0, i^*]) = \frac{i^*}{i_{max}} \text{ with } i^* \in [0, i_{max}])$.
- 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:
 $\pi_I(i) = \begin{cases} i \leq i^*: & i_{max} + m_I * i \\ i > i^*: & i_{max} + m_I * (2 * i^* - i) \end{cases} \text{ with } m_I > 0$,
- L : percentage extent of lying with: $L = \frac{a-i^*}{i^*}$ with $i^* > 0$,
- L_{guess} : percentage extent of suspected lying with: $L_{guess} = \frac{a-i_{guess}^*}{i_{guess}^*}$ with $i_{guess}^* > 0$,

⁴⁹ 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 my paper. However, in the results section, 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.

- \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 (A.2.1.) the advisor's set of best responses $S_{A;BR}$. Secondly, I will derive (A.2.2.) the investor's set of best responses $S_{I;BR}$. Thirdly, I will identify (A.2.3.) the set of game theoretical equilibria S_{equ} . Finally, I will outline (A.2.4.) some implications for the impact of both players' beliefs on their behavior in equilibrium.

A.2.1. The advisor's set of best responses $S_{A;BR}$

To begin with, the advisor aims to maximize their utility

$$\begin{aligned} U_A(a, i) &= \pi_A(i) - C_L(a) \\ &= m_A * i - C_L(a) \end{aligned} \tag{20}$$

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 ($a \geq i^* \leftrightarrow L \geq 0$).⁵⁰ 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} . It yields:

$$\forall a \in S_{A;BR}: a \in [i^*, i_{max}]. \quad (21)$$

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* $U_A^e(a, i_{guess})$:

$$\begin{aligned} U_A^e(a, i_{guess}) &= U_A(a, i = i_{guess}) \\ &= \pi_A(i = i_{guess}) - C_L(a) \\ &= m_A * i_{guess} - C_L(a) \end{aligned} \quad (22)$$

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:

$$\bar{T}_{guess} = \frac{i_{guess} - a}{a} \quad \leftrightarrow \quad i_{guess} = (\bar{T}_{guess} + 1) * a \quad (23)$$

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 $U_A^e(a, \bar{T}_{guess}(a))$ 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:

⁵⁰ 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 ($a > i^* \leftrightarrow L > 0$).

$$\begin{aligned}
U_A^e(a, \bar{T}_{guess}(a)) &= m_A * i_{guess} - C_L(a) \\
&= m_A * (\bar{T}_{guess}(a) + 1) * a - C_L(a)
\end{aligned} \tag{24}$$

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: $i = (\bar{T}_{guess} + 1) * a$) 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 $U_A^e(a, \bar{T}_{guess}(a))$ with respect to the advice $a \in [i^*, i_{max}]$.

It yields:

$$\begin{aligned}
\max_a \left(U_A^e(a, \bar{T}_{guess}(a)) \mid a \in [i^*, i_{max}] \right) &\rightarrow \frac{\partial U_A^e(a, \bar{T}_{guess}(a))}{\partial a} \stackrel{!}{=} 0 \\
&\leftrightarrow \frac{\partial (m_A * (\bar{T}_{guess}(a) + 1) * a - C_L(a))}{\partial a} = 0 \quad (25) \\
&\leftrightarrow \frac{\partial (C_L(a))}{\partial a} = m_A * \frac{\partial ((\bar{T}_{guess}(a) + 1) * a)}{\partial a}
\end{aligned}$$

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 $m_A * \frac{\partial ((\bar{T}_{guess}(a)+1)*a)}{\partial a}$. 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:

$$\forall a \in S_{A,BR}: \quad \frac{\partial (C_L(a))}{\partial a} = m_A * \frac{\partial ((\bar{T}_{guess}(a)+1)*a)}{\partial a} \quad \vee \quad a = i^* \vee a = i_{max}. \tag{26}$$

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

$$S_{A;BR} = \left\{ a^s \left| a^s, a^{s'} \in \left\{ a \in [i^*, i_{max}] \left| \frac{\partial(C_L(a))}{\partial a} = m_A * \frac{\partial((\bar{T}_{guess}(a) + 1) * a)}{\partial a} \right. \right. \right. \right. \\ \left. \left. \left. \vee a = i^* \vee a = i_{max} \right\} \wedge \forall a^{s'}: U_A^e(a^s) \geq U_A^e(a^{s'}) \right\} \quad (27)$$

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

A.2.2. The investor's set of best responses $S_{I;BR}$

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

$$U_I(a, i) = \pi_I(i) - C_{\bar{T}}(a, i) \\ = \begin{cases} i \leq i^*: & i_{max} + m_I * i - C_{\bar{T}}(a, i) \\ i > i^*: & i_{max} + m_I * (2 * i^* - i) - C_{\bar{T}}(a, i) \end{cases} \quad (28)$$

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^* ($a \geq i^* \leftrightarrow L \geq 0$).⁵¹ As the investor anticipates this ($a \geq i_{guess}^* \leftrightarrow L_{guess} \geq 0$), he or she always makes an investment i less or equal to their received advice a ($i \leq a \leftrightarrow \bar{T} \leq 0$).⁵² 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:

$$\forall i \in S_{I;BR}: i \in [i_{guess}^*, a]. \quad (29)$$

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

⁵¹ 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.

⁵² 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}^* :

$$\begin{aligned} U_I^e(a, i, i_{guess}^*) &= \pi_I(i)_{i^*=i_{guess}^*} - C_{\bar{T}}(a, i) \\ &= \begin{cases} i \leq i_{guess}^*: & i_{max} + m_I * i - C_{\bar{T}}(a, i) \\ i > i_{guess}^*: & i_{max} + m_I * (2 * i_{guess}^* - i) - C_{\bar{T}}(a, i) \end{cases} \end{aligned} \quad (30)$$

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:

$$L_{guess} = \frac{a - i_{guess}^*}{i_{guess}^*} \quad \leftrightarrow \quad i_{guess}^* = \frac{a}{L_{guess} + 1} \quad (31)$$

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} :

$$\begin{aligned} U_I^e(a, i, L_{guess}) &= U_I^e\left(a, i, i_{guess}^* = \frac{a}{L_{guess} + 1}\right) \\ &= \begin{cases} i \leq \frac{a}{L_{guess} + 1}: & i_{max} + m_I * i - C_{\bar{T}}(a, i) \\ i > \frac{a}{L_{guess} + 1}: & i_{max} + m_I * \left(2 * \frac{a}{L_{guess} + 1} - i\right) - C_{\bar{T}}(a, i) \end{cases} \quad (32) \\ &= \begin{cases} i < \frac{a}{L_{guess} + 1}: & i_{max} + m_I * i - C_{\bar{T}}(a, i) \\ i \geq \frac{a}{L_{guess} + 1}: & i_{max} + m_I * \left(2 * \frac{a}{L_{guess} + 1} - i\right) - C_{\bar{T}}(a, i) \end{cases} \end{aligned}$$

with $m_I > 0$ and $L_{guess} \geq 0$.⁵³

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

⁵³ 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 $i = i_{guess}^* = \frac{a}{L_{guess} + 1}$ it yields:

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

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:

$$\begin{aligned}
 \max_i \left(U_I^e(a, i, L_{guess}) \mid i \in \left[\frac{a}{L_{guess}+1}, a \right] \right) & \rightarrow \frac{\partial U_I^e(a, i, L_{guess})}{\partial i} \stackrel{!}{=} 0 \\
 & \Leftrightarrow -m_I - \frac{\partial (C_{\bar{T}}(a, i))}{\partial i} = 0 \quad (33) \\
 & \Leftrightarrow \frac{\partial (C_{\bar{T}}(a, i))}{\partial i} = -m_I
 \end{aligned}$$

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:

$$\forall i \in S_{I;BR}: \frac{\partial(C_{\bar{T}}(a,i))}{\partial i} = -m_I \vee i = \frac{a}{L_{guess}+1} \vee i = a. \quad (34)$$

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

$$S_{I;BR} = \left\{ i^s \mid i^s, i^{s'} \in \left\{ i \in \left[\frac{a}{L_{guess}+1}, a \right] \mid \frac{\partial(C_{\bar{T}}(a,i))}{\partial i} = -m_I \vee i = \frac{a}{L_{guess}+1} \vee i = a \right\} \right. \\ \left. \wedge \forall i^{s'}: U_I^e(i^s) \geq U_I^e(i^{s'}) \right\} \quad (35)$$

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

A.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:

$$S_{equ} = \left\{ (a^s, i^s) \mid \left(a^s, a^{s'} \in \left\{ a \in [i^*, i_{max}] \mid \frac{\partial(C_L(a))}{\partial a} = m_A * \frac{\partial(i^s(a))}{\partial a} \vee a = i^* \vee a = i_{max} \right\} \right. \right. \\ \left. \wedge \forall a^{s'}: U_A^e(a^s) \geq U_A^e(a^{s'}) \right) \\ \wedge \left(i^s, i^{s'} \in \left\{ i \in [i^*, a^s] \mid \frac{\partial(C_{\bar{T}}(a^s, i))}{\partial i} = -m_I \vee i = i^* \vee i = a^s \right\} \right. \\ \left. \wedge \forall i^{s'}: U_I^e(i^s) \geq U_I^e(i^{s'}) \right) \Bigg\} \quad (36)$$

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 A1.

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

Table A1. 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.

A.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 $C_T(a, i)$) 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 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. It yields the following conditions:

$$\frac{\partial}{\partial a} \left(\frac{\partial(C_L(a))}{\partial a} \right) = \begin{cases} a < i^*: & \leq 0 \\ a > i^*: & \leq 0 \end{cases} \quad (37)$$

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

$$\frac{\partial}{\partial i} \left(\frac{\partial(C_{\bar{T}}(a, i))}{\partial i} \right) = \begin{cases} i < a: & \leq 0 \\ i > a: & \leq 0 \end{cases} \quad (38)$$

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:

$$\frac{\partial(C_L(a))}{\partial a} = m_A * \frac{\partial(i(a))}{\partial a} \quad (39)$$

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 expects the absolute value of the investor's extent of (risk-reducing) mistrust ($|\bar{T}_{guess}(a)| \uparrow$ 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:

$ \bar{T}_{guess}(a) \uparrow$ (for all advice a)	$\rightarrow \bar{T}_{guess}(a) \downarrow$	(since: $\bar{T}_{guess}(a) \leq 0$)
	$\rightarrow i_{guess}(a) \downarrow$	(since: $i_{guess}(a) = (\bar{T}_{guess}(a) + 1) * a$)
	$\rightarrow \frac{\partial(i_{guess}(a))}{\partial a} \downarrow$	(since: $i_{guess}(a)$ becomes zero for $a = 0$, is concave, and increases monotonously. ⁵⁴ 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.)
	$\rightarrow \frac{\partial(i(a))}{\partial a} \downarrow$	(since: $i_{guess} = i$ due to correct beliefs in equilibrium)
	$\rightarrow \left(m_A * \frac{\partial(i(a))}{\partial a}\right) \downarrow$	(since: $m_A > 0$)
	$\rightarrow \frac{\partial(C_L(a))}{\partial a} \downarrow$	(since: $\frac{\partial(C_L(a))}{\partial a} = m_A * \frac{\partial(i(a))}{\partial a}$ must be fulfilled for any interior solution to the advisor's maximization problem in equilibrium)
	$\rightarrow C_L(a)_{\text{accepted by advisor}} \uparrow$	(since: $C_L(a)$ is concave and increases monotonously for $a \geq i^*$)
	$\rightarrow a \uparrow$	(since: $C_L(a)$ increases monotonously for $a \geq i^*$)
	$\rightarrow (a - i^*) \uparrow$	(since: i^* is constant)
	$\rightarrow L \uparrow$	(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 $|\bar{T}_{guess}|$) make the advisor engage in more dishonest behavior (i.e., a larger extent of lying L).

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

$$\frac{\partial(C_{\bar{T}}(a, i))}{\partial i} = -m_I \quad (40)$$

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:

⁵⁴ 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 .

$$\begin{aligned}
m_I \uparrow &\rightarrow (-m_I) \downarrow \\
&\rightarrow \frac{\partial(C_{\bar{T}}(a, i))}{\partial i} \downarrow && \text{(since: } \frac{\partial(C_{\bar{T}}(a, i))}{\partial i} = -m_I \text{ must be fulfilled for any interior solution to the investor's maximization problem in equilibrium)} \\
&\rightarrow C_{\bar{T}}(a, i)_{\text{accepted by investor}} \uparrow && \text{(since: } C_{\bar{T}}(a, i) \text{ is concave and decreases monotonously for } i \leq a) \\
&\rightarrow i \downarrow && \text{(since: } C_{\bar{T}}(a, i) \text{ decreases monotonously for } i \leq a).
\end{aligned}$$

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:

$$i_{guess}^* = \frac{a}{L_{guess} + 1} \quad (41)$$

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:

$$\begin{aligned}
L_{guess} \uparrow &\rightarrow \left(\frac{a}{L_{guess} + 1} \right) \downarrow && \text{(since: } a \geq 0 \text{ and } L_{guess} \geq 0) \\
&\rightarrow i_{guess}^* \downarrow && \text{(since: } i_{guess}^* = \frac{a}{L_{guess} + 1}) \\
&\rightarrow \min(i) \downarrow && \text{(since: } i \in [i_{guess}^*, a]) \\
&\rightarrow \max(|\bar{T}|) \uparrow && \text{(since: } i \leq a \leftrightarrow \bar{T} \leq 0 \text{ and } \bar{T} = \frac{i-a}{a}).
\end{aligned}$$


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.


Appendix B. Materials

B.1. Instructions for the advisor (for the Continuous Deception Game)

The following is a translation of the instructions that I presented to the advisors in *classEx*.⁵⁵ Original German instructions are available upon request.

B.1.1. Instructions before the experiment started (for the advisor)



 **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 B1. Instructions for the advisor – Part 1/5

⁵⁵ Technical instructions regarding the use of *classEx* on mobile phones were presented separately from these instructions and are omitted here.

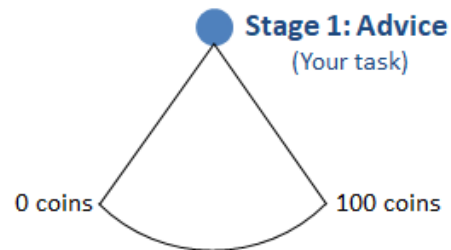
Procedure within a round:

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)

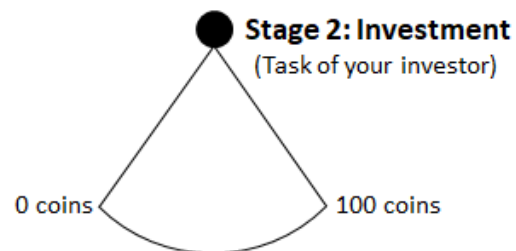


Figure B2. Instructions for the advisor – Part 2/5

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.

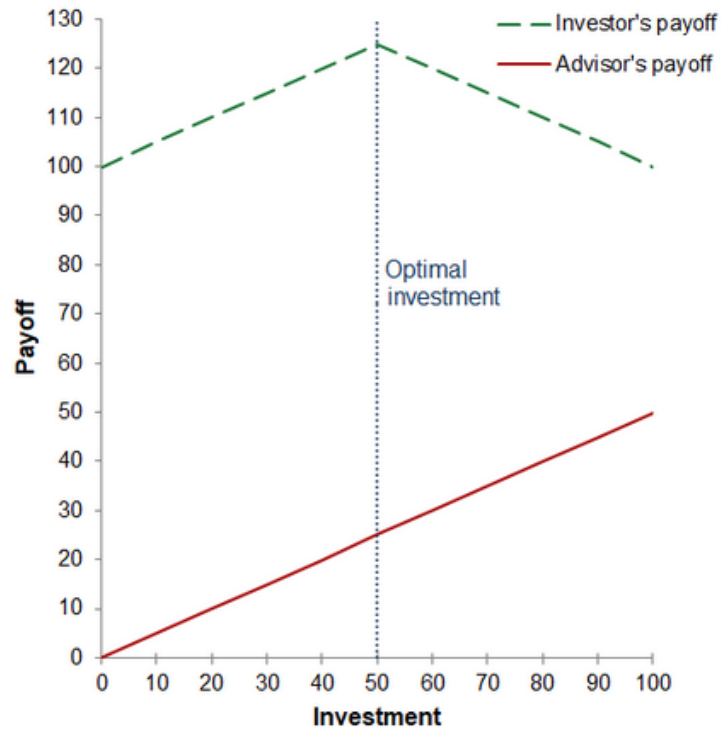
Optimal investment:

- **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 B3. 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 B4. 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 B5. Instructions for the advisor – Part 5/5

B.1.2. Input screen for one single round of the Continuous Deception Game (for the advisor)

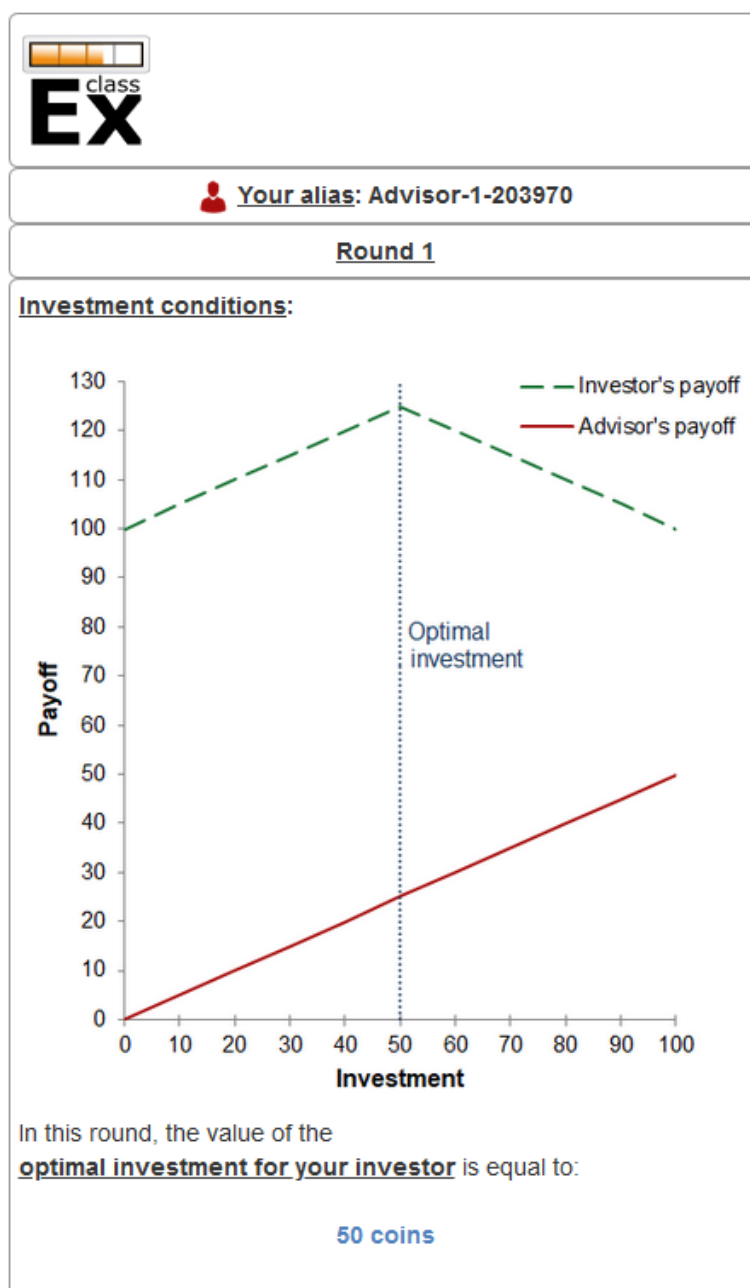


Figure B6. Input screen for the advisor – Part 1/2

Please decide on:

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

2.) Expected amount of investment:
Estimate the amount of the investment that you expect your investor to make after receiving your advice.


(Please enter any value between 0 and 100 coins in the respective input fields.)

1.) Your advice to the investor:


coins

2.) Expected amount of investment:

coins

 **Your payoff in this round:**
(if the investor exactly follows your advice)

coins

 **Investor's payoff in this round**
(if the investor exactly follows your advice)

coins

Send


Figure B7. Input screen for the advisor – Part 2/2


Note. The two greyed-out fields in Figure B7 automatically indicate the payoffs that both players would receive if the investor exactly follows the advice number that the advisor has entered in the first input field.

B.2. Instructions for the investor (for the Continuous Deception Game)

The following is a translation of the instructions that I presented to the investors in *classEx*.⁵⁶ Original German instructions are available upon request.

B.2.1. Instructions before the experiment started (for the investor)



 **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 B8. Instructions for the investor – Part 1/5

⁵⁶ Technical instructions regarding the use of *classEx* on mobile phones were presented separately from these instructions and are omitted here.

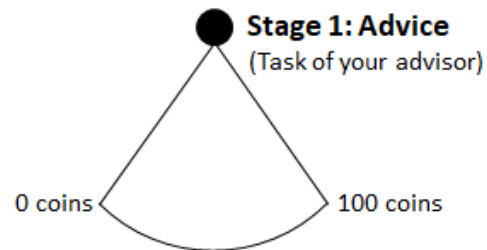
Procedure within a round:

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)

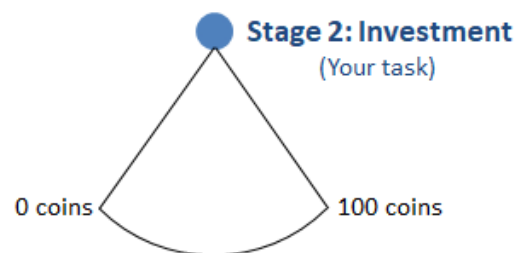


Figure B9. Instructions for the investor – Part 2/5

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.

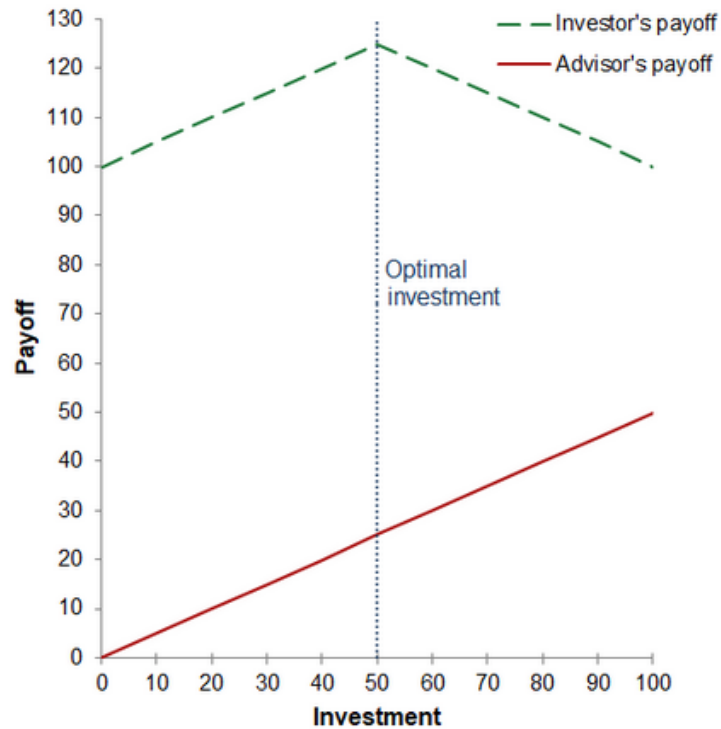
Optimal investment:

- **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 B10. 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 B11. 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 B12. Instructions for the investor – Part 5/5

B.2.2. Input screen for one single round of the Continuous Deception Game (for the investor)

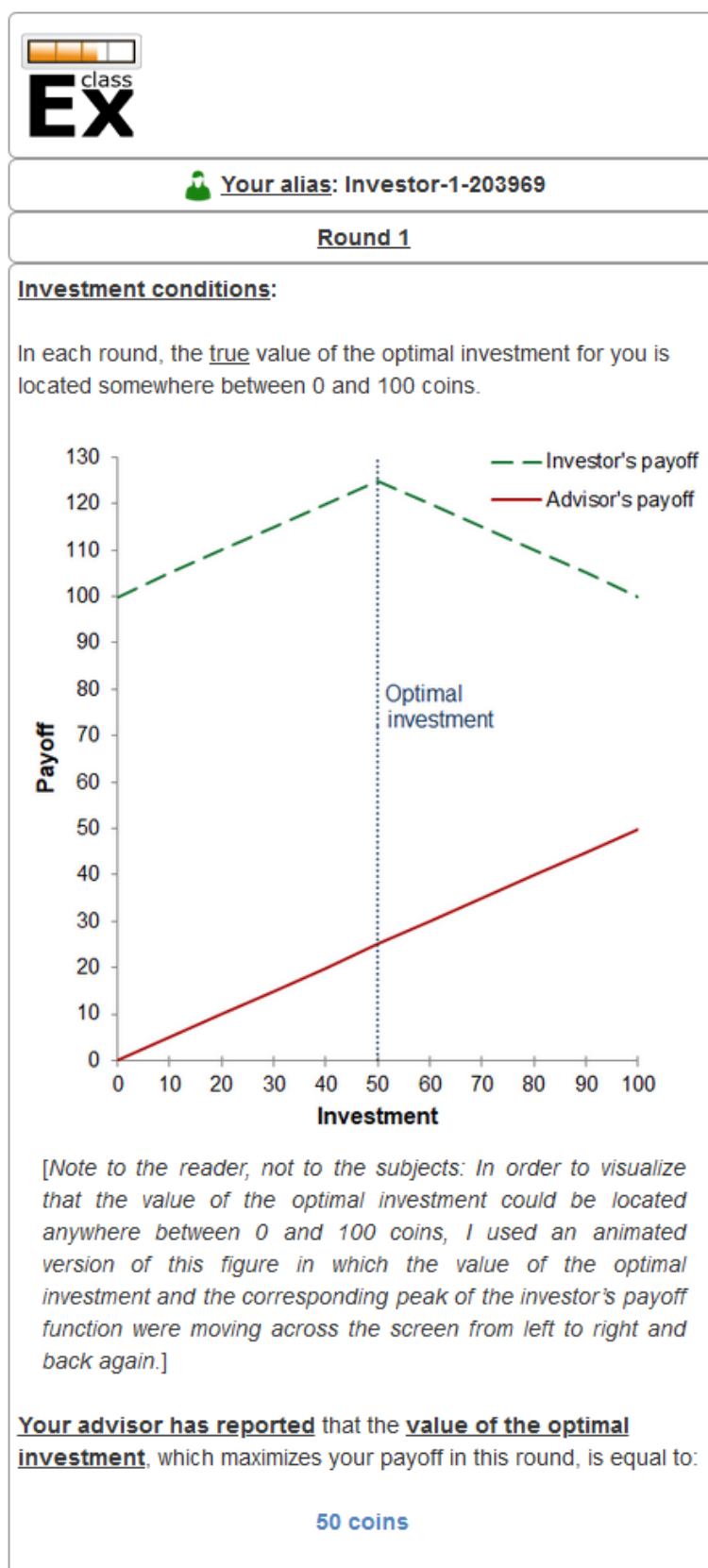


Figure B13. Input screen for the investor – Part 1/2

Please decide on:

1.) Your amount of investment:
Decide how many coins you want to invest in this round.

2.) Estimated value of the optimal investment:
Estimate which amount of investment maximizes your payoff in this round.

(Please enter any value between 0 and 100 coins in the respective input fields.)

1.) Your amount of investment:

coins

2.) Estimated value of the optimal investment:

coins

Send

Figure B14. Input screen for the investor – Part 2/2

Appendix C. 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 (C.1.) with their self-assessed preference for risk and honesty. Then, I will examine how consistent the observed behavior of the investors (C.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).

C.1. 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 ones. Interestingly, they rated their own preference for risk as moderately but barely non-significantly higher than they rated the one 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 C1. *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 ($|L|$)⁵⁷ 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 C2. *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.*

C.2. 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 to the group. Moreover, the investors' self-assessed preference for risk correlates significantly positively

⁵⁷ 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.

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: $\rho = 0.552$ 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 C3. *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 ($|\bar{T}|$)⁵⁸ on average over all ten rounds (Spearman's rank correlation with outlier-cleaned values: $\rho = -0.554$ 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: $\rho = 0.883$ 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 C4. *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.*

⁵⁸ Here, I use the absolute value of the percentage extent of mistrust ($|\bar{T}|$), since 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.

Appendix D. Temporal consistency of players' 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 (D.1.) and then continue with the investors (D.2.).

D.1. Temporal consistency of advisor decisions

Figure D1 visualizes lag plots for the percentage extent of lying (L) on the left (D1a) and the percentage extent of expected mistrust (\bar{T}_{guess}) on the right (D1b).⁵⁹ 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 D1a) are lagged by five rounds. As for the plot on the right (Figure D1b), 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 D1b) 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.⁶⁰

The lag plot for the percentage extent of lying (L) in Figure D1a 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.⁶¹

Finding D1. *The advisors' lying behavior (L) was mostly consistent over time.*

⁵⁹ For the purpose of illustration, only the most relevant section of the plot in Figure D1a is displayed.

⁶⁰ 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 $p = 0.250$).

⁶¹ This refers to whether they gave honest advice, lied by understating, or lied by overstating.

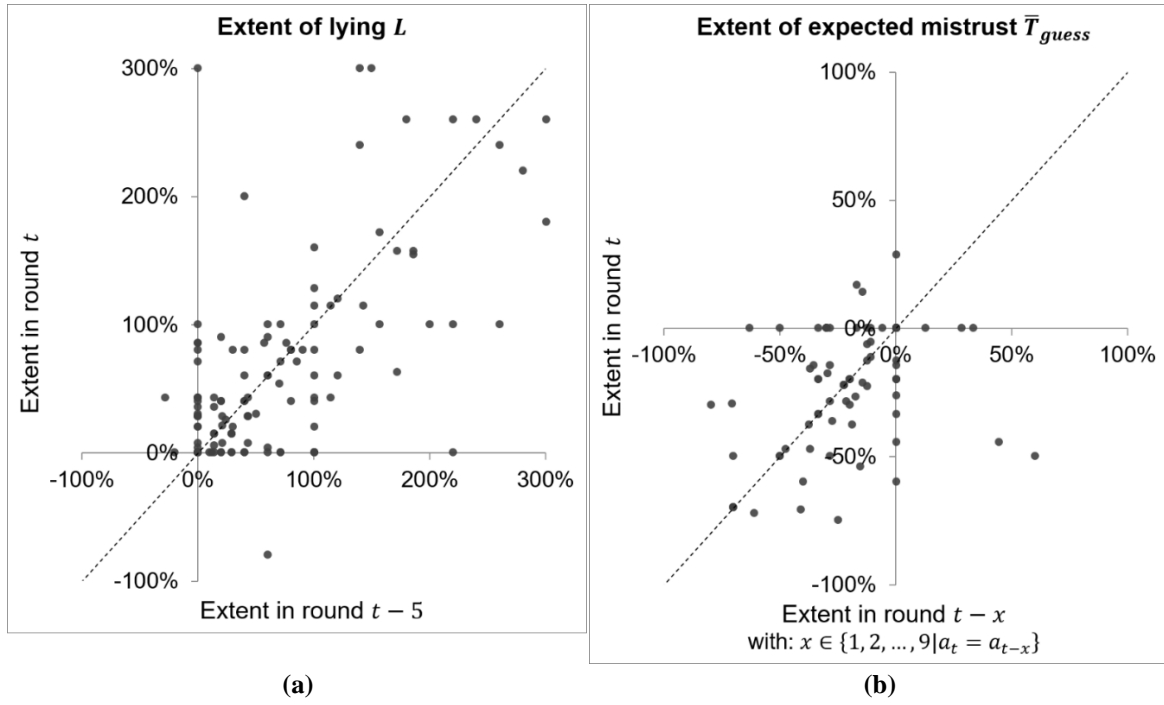


Figure D1. Temporal consistency of the advisors' behavior and first-order beliefs: **(a)** Lag plot of the extent of lying L ; **(b)** Lag plot of the extent of expected mistrust \bar{T}_{guess}

Turning to the lag plot for the percentage extent of expected mistrust (\bar{T}_{guess}) in Figure D1b, 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.⁶²

Finding D2. *The advisors' first-order beliefs about their investors' mistrust (\bar{T}_{guess}) were mostly consistent over time.*

⁶² This refers to whether they expected trusting behavior, risk-reducing mistrust, or risk-seeking mistrust from their investors.

D.2. Temporal consistency of investor decisions

Figure D2 displays lag plots for the percentage extent of mistrust (\bar{T}) on the left (D2a) and the percentage extent of suspected lying (L_{guess}) on the right (D2b).⁶³ 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 D2 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.⁶⁴

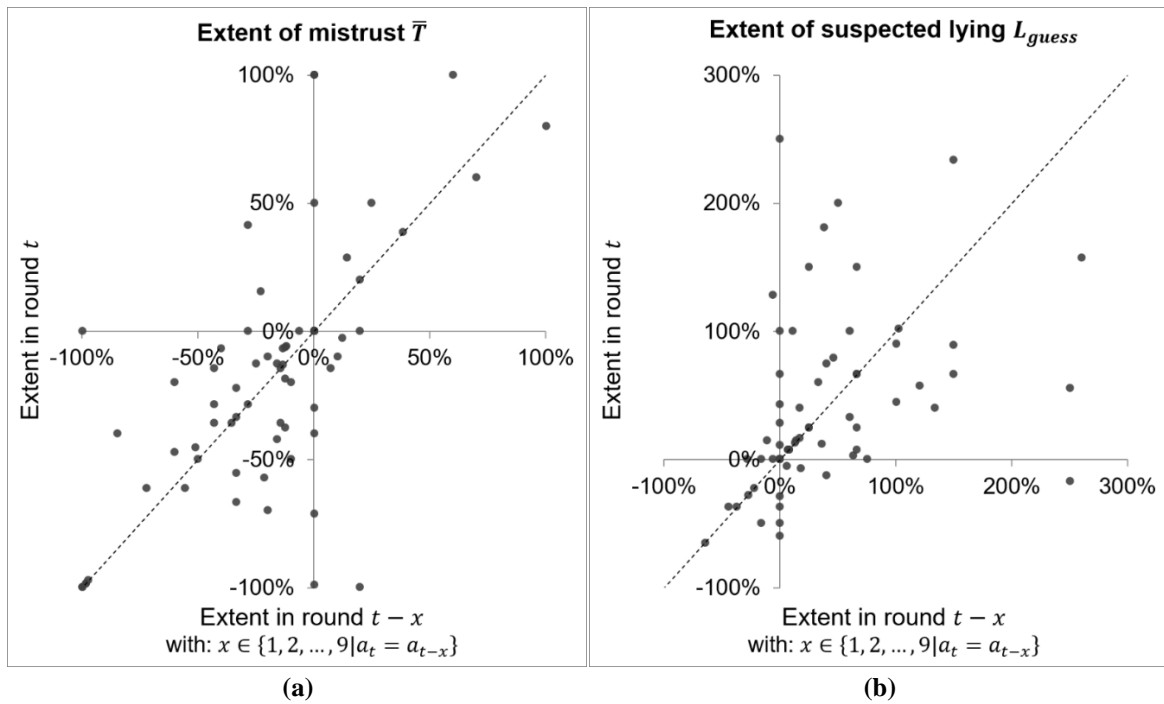


Figure D2. Temporal consistency of the investors' behavior and first-order beliefs: (a) Lag plot of the extent of mistrust \bar{T} ; (b) Lag plot of the extent of suspected lying L_{guess}

The lag plot for the percentage extent of mistrust (\bar{T}) in Figure D2a 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 plot shows several different trends in the development of the investors' (mis)trust over time: Firstly,

⁶³ For the purpose of illustration, only the most relevant section of the plot in Figure D2b is displayed.

⁶⁴ 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 $p = 0.899$) *nor* with the change in the extent of suspected lying over the time lag (Spearman's rank correlation: $\rho = -0.107$ with $p = 0.358$). 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.

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 they mistrusted it to a different extent. Overall, in 77.63% of cases, the investors' mistrusting behavior had the same orientation before and after the time lag.⁶⁵

Finding D3. *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 D2b that the investors' first-order beliefs about their advisors' lying behavior were also generally consistent over time. This is due to 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.⁶⁶

Finding D4. *The investors' first-order beliefs about their advisors' lying behavior (L_{guess}) were mostly consistent over time.*

⁶⁵ This refers to whether they followed their received advice, engaged in risk-reducing mistrust, or engaged in risk-seeking mistrust.

⁶⁶ This refers to whether they expected their advisors to tell the truth, to lie by understating, or to lie by overstating.