MAGKS

Joint Discussion Paper Series in Economics

by the Universities of

Aachen · Gießen · Göttingen
Kassel · Marburg · Siegen

ISSN 1867-3678

No. 30-2020

Tobias Beck

Lying and Mistrust in the Continuous Deception Game

Coordination: Bernd Hayo • Philipps-University Marburg
School of Business and Economics • Universitätsstraße 24, D-35032 Marburg
Tel: +49-6421-2823091, Fax: +49-6421-2823088, e-mail: hayo@wiwi.uni-marburg.de

Lying and Mistrust
in the Continuous Deception Game

By Tobias Beck[*]

June 2020

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. The extents of lying and mistrust are both measured on continuous scales. This 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%. The senders seldomly lie to the fullest possible extent as they correctly expect the receivers to disproportionally mistrust lies of such a high extent. This indicates that people make strategic considerations about their potential to manipulate others when lying. In line with this, I discover that lying and mistrusting behavior can be predicted by first-order beliefs about the other player. Consistent with previous studies, my findings support the conjecture that lying costs increase with the extent of lying. In addition, I provide evidence for some endogenous preference for trust. Both players' behaviors and beliefs are consistent over time. Moreover, my ex ante classification of both players' strategy sets is consistent with their ex post self-assessment of their own behavior within the experiment.

Keywords Lies, Honesty, Mistrust, Deception Game, Investments, Asymmetric information, Experimental Design

JEL Classification C91, D01, D82, G41

[*] Corresponding author: Institute of Economics, University of Kassel, Nora Platiel-Straße 4, 34127 Kassel, Germany (email: tobi.beck@gmx.net; ORCID: 0000-0002-4162-3648)

1. Introduction

The measurement of (dis)honesty and (mis)trust is a challenge for researchers in various fields of study. In particular, in experimental economics, there exists 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 extents of lying *and* mistrusting behavior 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. 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 [1]. 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 the Madoff investment scandal, which was discovered in 2008 and resulted in the SEC charging former investment advisor Bernard Madoff with securities fraud for committing an $18 billion Ponzi scheme [2]. For years, Madoff had deceived certain investors by paying them returns out of the principal received from other, different investors, without even investing their money in real assets. Even if real investments are made, the list of investment advisors who gave misleading or false advice to their clients to invest in unprofitable assets is long (e.g., [3-5]).

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, 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 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* is inspired by Gneezy [6] as well as Erat and Gneezy [7]. 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]

[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.

I will 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 Continuous Deception Game. This will allow me to define several key variables to measure lying, mistrust, and players' expectations of each other in the game. Based on that, I will categorize all feasible strategies in the game and show which of them are rational from a game theoretical perspective. I will proceed by explaining the precise implementation of my experiment by discussing the design in section 4 and the experimental procedures in section 5. Due to the novelty of my experimental design, the hypotheses derivation in section 6 aims to cover a wide range of theoretical considerations to formulate specific expectations towards both players' behavior and their beliefs in the game. I will report the results in section 7. This section is divided into five subsections: (*7.1.*) the main results on both players' overall behavior and first-order beliefs, (*7.2.*) an analysis of the sensitivity of honesty, trust, and respective beliefs to in-game variables, (*7.3.*) the testing of both players' behavior and first-order beliefs for temporal consistency, (*7.4.*) an analysis of both players' strategic orientation based on the relationship between their behavior and beliefs, and finally (*7.5.*) the testing of the consistency of my ex ante classification of both players' strategies with their ex post self-assessment of their own behavior. This range of results is intended to inspire new applications of my novel experimental design. In addition, I offer valuable insights into the mechanisms of lying and mistrust in strategic deception in the face of a task with contradicting incentives. Section 8 summarizes and discusses my results in the light of the existing literature. I will conclude my most important findings in section 9.

2. Literature review

2.1. Honesty

There is broad evidence that people have some form of preference for honest behavior (e.g., [6-15]). In general, it is argued that this preference is intrinsic rather than extrinsic. In support of this, Pruckner and Sausgruber [16] show that appealing to honesty can mitigate dishonest behavior more effectively than a reminder of legal norms. Often this is at least partially explained with the concept of lie aversion (e.g., [6,9-13]). Vanberg [15] even provides experimental evidence that people dislike the act of lying per se. Another approach comes from Charness and Dufwenberg [8] who show that people's preference for honest behavior in their experiment can be explained by guilt aversion. On that subject, Battigalli et al. [17] argue that, in some situations, guilt can provide a psycho-foundation for honesty. Moreover, Gneezy et al. [10] 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 [18] 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" [18] (p. 193, italics in original). In line with this, Mazar et al. [14] find that directing one's attention to moral standards can lower one's tolerance for dishonest behavior. Furthermore, Mazar et al. [14] 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 [19].

In many studies these internal motivators for preferences for honesty are modeled by the concept of *costs of lying* (e.g., [10,13,20]). 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. [13] find that the aversion to lying increases with the size of the lie. Following this idea, Beck et al. [20] 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 [9] dice experiment. In addition, Gneezy et al. [10] 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. These costs 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 being that social trust can raise economic growth rates [21-23]. Trust is important for the banking sector in particular [24]. With respect to the individual, Barefoot et al. [25] show that high levels of trust on Rotter's [26] interpersonal trust scale are associated with better self-rated health and more life satisfaction. In addition, Kuroki [27] finds, by analyzing survey data, that interpersonal trust has positive and significant effects on individual happiness. In line with this, Gurtman [28] 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. [29] 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 [30] (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 [30] 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. "Trust" here refers to *honesty-related trusting behavior*, i.e., one's reliance on another person's honesty. "Trusting beliefs" in this paper are examined in relation to 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 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., [8,31-33]). 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 expectations of social cooperation, but not of honest behavior. 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. This trust, however, is in conflict with their own 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 [24,34]. The same applies for advisors who have more information than the investors they consult. In general, insiders have superior information to investors [35]. Hence, insiders might use their informational power to manipulate investors to invest in their firm [34]. 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 [34], 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

[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 [36].

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 [34], 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 [6], Sánchez-Pagés and Vorsatz [37], Dreber and Johannesson [38], Sutter [39], Erat and Gneezy [7], López-Pérez and Spiegelman [12], Peeters et al. [40], Peeters et al. [41], Jacquemet et al. [42], Jacquemet et al. [43], Vranceanu and Dubart [44], and Gneezy et al. [45] 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 [7] *Deception Game* (which is a sender-receiver game that originated from Gneezy [6]), I will discuss their experimental design in more 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 roll of the die, each stating that "the outcome from the roll of the 6-sided die is…" [7] (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 [7] 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 [7] 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 [7] experiment, both players' decisions can be considered as *binary-like*[4] choices. Therefore, 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 [7] paper in a good way, in reality the ranges of dishonesty and mistrust are more continuous.

Lundquist et al. [13] 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 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

[3] A similar distinction is also used by Gneezy [6].

[4] Obviously, 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 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 be indifferent between the five untruthful messages that he or she could send, too.

[5] Translated to Gneezy et al.'s [10] model of intrinsic costs of lying, Erat and Gneezy's [7] 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.

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.[8] These frequencies do not represent feasible strategies in the game but proportions of players 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.[9] As a consequence, binary (or even discrete) choices do not allow measuring the real extent to which the sender (or, respectively, the receiver) would prefer to lie (or, respectively, to mistrust) when given the chance.

To the best of my knowledge, there does not exist a single experiment that allows examining honesty *and* trust as continuous variables on comparable scales. This is where this paper intends 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 defined by Sobel [34]) that is framed as an investment game.[10] Therefore, it features an advisor with complete and an investor with incomplete information. This novel game is largely inspired by the experimental design of Gneezy [6] and Erat and Gneezy [7]. It expands Erat and Gneezy's [7] 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*.

[6] Lundquist et al. [13] 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 [10] model of intrinsic costs of lying, Lundquist et al.'s [13] 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] Another approach is used by Vranceanu and Dubart [44] who modify Gneezy's [6] sender-receiver game by letting the senders repeatedly choose between lying and truth-telling while varying the senders' payoffs that they can gain from misleading the receivers. By this means, Vranceanu and Dubart [44] can identify the senders' individual reservation payoffs, i.e., their individual thresholds for switching from faithful to deceitful communication. While this is a clever design, which allows observing preferences for deception on an individual level, it does not capture the size of the lie that the players actually prefer.

[9] Of course, 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 [9], which allows distinguishing partial from payoff-maximizing lies. Another design is the experimental design of Gneezy et al. [10] 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.

[10] The idea of framing my experiment as an investment game is inspired by Berg et al. [46] who designed an investment game in order to introduce continuous variables into the Trust Game.

3.1. Payoff structure and incentives

Towards the end of the Continuous Deception Game, 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. It is randomly determined by nature before the game starts.[11] Both players have different payoff functions:

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

$$\pi_A(i) = m_A * i \tag{1}$$

with $m_A > 0$.

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 $i^* \neq i_{max}$: $i > i^*$).

On the other hand, 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}. Consequently, if the investor later 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 $\pi_I(i)$ will be defined by the following split function of the investment i, which decreases linearly by deviating downward or upward from the optimal investment i^*:

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

with $m_I > 0$.

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 starting 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.

---

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

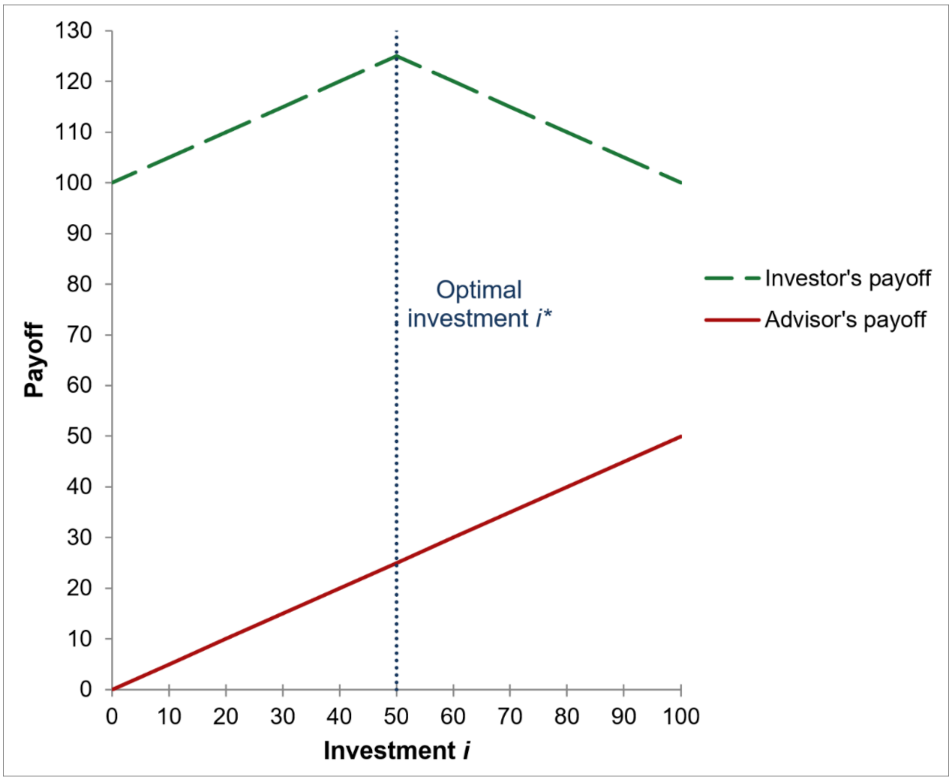


Figure 1. Payoff structure of the Continuous Deception Game.

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.

3.2. The game process

Figure 2 summarizes the process of the Continuous Deception Game. The game itself consists of two main stages. As mentioned before, the optimal investment i^* is randomly and in secret determined 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 truthfully 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.[12] 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

[12] This is inspired by Sutter [39] 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. Just as Sutter [39], I refrain from monetarily incentivizing this guess, since “there is evidence that eliciting expectations with or without monetary rewards for accuracy does not yield significantly different results” (Sutter [39], p. 50, referring to Camerer and Hogarth [47]).

investment i^*.[13] 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.

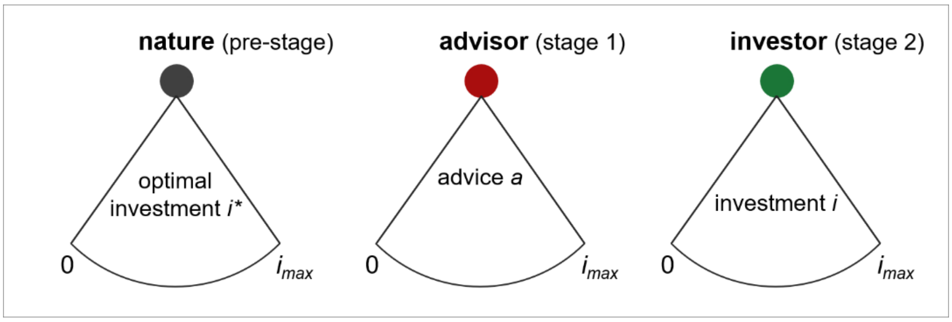


Figure 2. Game tree of the Continuous Deception Game.

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.

3.3. Key variables

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

3.3.1. (Suspected) lying

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

$$L = \frac{a - i^*}{i^*} \tag{3}$$

with $i^* > 0$.[14]

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:

$$L_{guess} = \frac{a - i^*_{guess}}{i^*_{guess}} \tag{4}$$

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

Note: 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 $a \neq i^*_{guess}$: $L_{guess} \neq 0$). In particular, guessing

[13] 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 [39,47].

[14] 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 percentage extent of lying could 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.

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 does not follow the received advice a is defined as:

$$\bar{T} = \frac{i - a}{a} \tag{5}$$

with $a > 0$.

Note: 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 $i < a$: $\bar{T} < 0$), whereas making an investment i above the advice a shall be defined as *risk-seeking mistrust* (if $i > a$: $\bar{T} > 0$), since in this game higher investments i are associated with a higher risk.[15]

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

$$\bar{T}_{guess} = \frac{i_{guess} - a}{a} \tag{6}$$

with $a > 0$.

Note: 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 $i_{guess} \neq a$: $\bar{T}_{guess} \neq 0$). In particular, guessing that the investment i_{guess} is below the given advice a shall be defined as *expecting risk-reducing mistrust* (if $i_{guess} < a$: $\bar{T}_{guess} < 0$), whereas guessing that the investment i_{guess} is above the advice a shall be considered *expecting risk-seeking mistrust* (if $i_{guess} > a$: $\bar{T}_{guess} > 0$), since in this game higher investments i are associated with a higher risk.

3.3.3. (Expected) misinvestment

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

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

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 $i_{guess} \neq i^*$: $F_{A;guess} \neq 0$). In particular, guessing that the investment i_{guess} is below the true optimum i^* can be defined as *expecting an underinvestment* (if $i_{guess} < i^*$: $F_{A;guess} < 0$), whereas guessing that the investment i_{guess} is above the optimal investment i^* shall be considered as *expecting an overinvestment* (if $i_{guess} > i^*$: $F_{A;guess} > 0$).

[15] The distinction between risk-reducing and risk-seeking mistrust is in line with Sapienza et al. [29] who find that mistrust is associated with a preference for risk aversion in the Trust Game. In addition, the variance of the investor's payoff increases with the value of the investment, which also speaks in favor of this terminology.

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

$$F_{I;guess} = \frac{i - i^*_{guess}}{i^*_{guess}} \tag{8}$$

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

$$F = \frac{i - i^*}{i^*} \tag{9}$$

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 $i \neq i^*$: $F \neq 0$). In particular, 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 measuring 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 Continuous Deception Game, 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.

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 [7] who use the expected change in payoffs in order to distinguish different types of lies,[16] I will 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 $F_{A;guess}$).

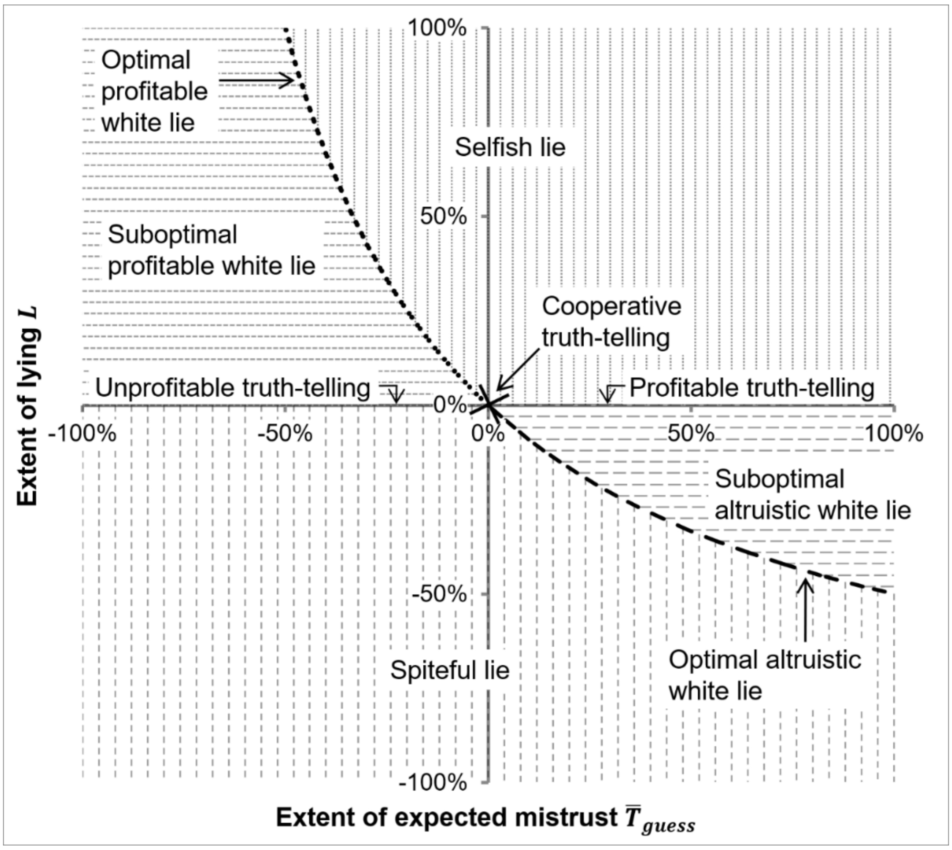


Figure 3. Taxonomy of lies and truth-telling for the advisor.

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 ($F_{A;guess} = 0$).

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 ($F_{A;guess} < 0$).

[16] As described above, Erat and Gneezy [7] 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 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.

Table 1. Definition of classes of advisor strategies.

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

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 will 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.

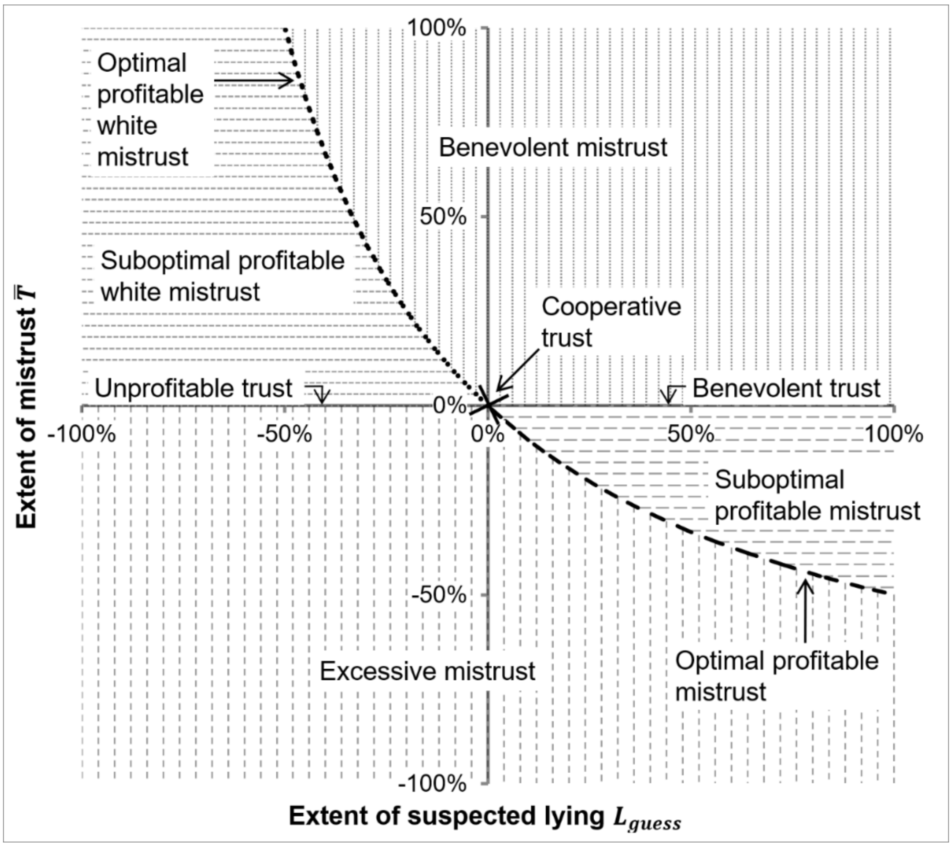


Figure 4. Taxonomy of mistrust and trust for the investor.

As summarized in Table 2, there are the nine following 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 ($F_{I;guess} = 0$).

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.

Table 2. Definition of classes of investor strategies.

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

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. Therefore, the next subsection is dedicated to find rational strategies that consider individual preferences for honesty and trust as well as personal beliefs about the other player.

3.5. Game theoretical perspective

In this subsection, I intend to solve the Continuous Deception Game by identifying its set of game theoretical equilibria and thereby determine strategies that are more likely to be picked by rational players. My game is a sequential game with incomplete information.[17] For this approach, both players are modeled as risk-neutral rational players who are seeking to maximize their expected utility, given their beliefs about the other player.

3.5.1. Only monetary motivation

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

This type of investor has no reason to trust the advice a and therefore disregards this information. As a result, he or she assumes the optimal investment i^* to be any random number between 0 and the maximal investment i_{max} with equal probability ($P([0, i^*]) = \frac{i^*}{i_{max}}$ with $i^* \in [0, i_{max}]$). In this case, the advice a has absolutely no impact. Based on that, the investor's expected payoff $\pi^e_{I;hoec}(i)$ can be described as the following function of the investment i:

[17] 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 [48-50]).

$$\pi^e_{I;hoec}(i) = i_{max} + m_I * \left(i - \frac{i^2}{i_{max}}\right) \tag{10}$$

with $m_I > 0$.[18]

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

$$i_{BR;hoec} = \frac{i_{max}}{2}. \tag{11}$$[19]

This investment results in the following expected payoff $\pi^e_{I;BR;hoec}$:

$$\pi^e_{I;BR;hoec} = \pi^e_{I;hoec}(i = i_{BR;hoec}) = i_{max} * \left(1 + \frac{m_I}{4}\right) \tag{12}$$

with $m_I > 0$.

In short, this 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, the *homo oeconomicus* 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 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:

$$S_{A;BR;hoec} = \{a | a \in [0, i_{max}]\}. \tag{13}$$

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

$$\pi^e_{A;BR;hoec} = \pi_A(i = i_{BR;hoec}) = m_A * i_{BR;hoec} = i_{max} * \frac{m_A}{2} \tag{14}$$

with $m_A > 0$.

Since all advice a will result in this same expected payoff $\pi^e_{A;BR;hoec}$, 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 Continuous Deception Game can be defined as:

$$S_{equ;hoec} = \left\{(a^s_{hoec}, i^s_{hoec}) \middle| a^s_{hoec} \in [0, i_{max}] \wedge i^s_{hoec} = \frac{i_{max}}{2}\right\}. \tag{15}$$

3.5.2. Monetary and non-monetary motivation

Suppose now that both players care for more than their monetary payoffs. Since the Continuous Deception Game 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., [8,14,15,17]). Since there exist 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 [25,27], while mistrust can lead to negative ones [28]. For that reason, I assume that players have

[18] The function $\pi^e_{I;hoec}(i)$ constitutes the expected value of the investor's payoff $\pi_I(i)$ as a function of the investment i. For a detailed derivation of this function, see appendix A.1.

[19] For a detailed derivation of the investor's best response $i_{BR;hoec}$ and the corresponding expected payoff $\pi^e_{I;BR;hoec}$, see appendix A.2.

a preference for trust. Now, in order to specify these *homines morales* [19], 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^*.[20] However, the nature of this function is specified by the advisor's type.[21] As a result, the advisor's utility $U_A(a,i)$ can be defined as the following function of their given advice a and the investment i:

$$U_A(a,i) = \pi_A(i) - C_L(a). \tag{16}$$

Analogously, 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.[22] Based on this, the investor's utility $U_I(a,i)$ can be defined as the following function of the advice a and their investment i:

$$U_I(a,i) = \pi_I(i) - C_{\bar{T}}(a,i). \tag{17}$$

[20] This is in line with Lundquist et al. [13] 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 my experiment measures a combination of Gneezy et al.'s [10] 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, the advisor's expectation towards their own payoff should be strongly associated with their given advice, since their payoff is designed as a linear function of the investment. 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. [10], 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. The investor has no way to know for sure if the advice is a lie. However, as the advisor has a monetary incentive to advise an excessive investment, it is reasonable to suppose that, the higher the advice, the higher is the likelihood that the investor perceives it as dishonest. For this reason, and even though each value of the true optimal investment can come up with the same probability, the advisor's concerns about how he or she is perceived by the investor should increase with the extent of lying by overstating. 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.

[21] More precisely, 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 $\left(|L| = \left|\frac{a-i^*}{i^*}\right|\right)$: $\frac{\partial(C_L(a))}{\partial a} = \begin{cases} a < i^*: & \leq 0 \\ a > i^*: & \geq 0 \end{cases}$.

[22] To specify, the investor's costs of mistrust $C_{\bar{T}}(a,i)$…

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

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

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

Also note that one 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.[23]

The *advisor's set of best responses* $S_{A;BR}$ depends on their type and, thus, on their moral costs of lying $C_L(a)$. These costs incentivize giving advice a equal to the true optimal investment i^* ($a = i^*$). In general, an honest type of advisor incurs higher moral costs when lying than when a dishonest advisor lies. This means that, on the one hand, a completely honest advisor would suffer tremendously from lying and, therefore, in all probability would give truthful advice ($a = i^* \leftrightarrow L = 0$). On the other hand, a completely dishonest advisor would not suffer from lying and, thus, would try to maximize solely their monetary payoff $\pi_A(i)$ by giving advice a that deviates from the optimal investment i^* if necessary ($a \neq i^* \leftrightarrow L \neq 0$). In summary, this means that the advisor remains with a monetary incentive to try and get the investor to make the highest possible investment ($i = i_{max}$), while he or she – depending on their type – now also has a non-monetary, rather intrinsic, incentive to give honest advice ($a = i^* \leftrightarrow L = 0$).

Aware of the possibility that the advisor wants to avoid lying, the investor now can use the advice a as a reference point for their investment i. Here, the *investor's set of best responses* $S_{I;BR}$ also depends on their type and, therefore, on their costs of mistrust $C_{\bar{T}}(a, i)$. The more trusting the investor, the more he or she follows the advice a. Therefore, a completely trusting investor would exactly follow the advice a ($i = a \leftrightarrow \bar{T} = 0$), while a completely mistrusting one would try to maximize their monetary payoff $\pi_I(i)$, most likely resulting in an investment i unequal to the advice a ($i \neq a \leftrightarrow \bar{T} \neq 0$).

The prior probability distribution over all of these possible realizations of investor types is in turn known to the advisor. Thus, he or she now has a chance to have influence on the investment i by giving strategic advice a. In this way, the advisor can try to impact their own monetary payoff $\pi_A(i)$.[24] Again, this is common knowledge to both players. As a result, both players know that there is a chance that the advisor could give honest advice ($a = i^* \leftrightarrow L = 0$), while there is also a chance that the investor could rely on it ($i = a \leftrightarrow \bar{T} = 0$). However, this depends on both players' types. Since the other's type is not known to the players, their decisions also depend on their beliefs about each other's type. Here, their beliefs are about the nature of the other's cost function $C_L(a)$ or $C_{\bar{T}}(a, i)$, respectively. With that, both players' sets of best responses ($S_{A;BR}$ and $S_{I;BR}$) and, thus ultimately, their behaviors in the game (L and $\bar{T}$) are determined conjointly by their own type and their beliefs about the other player. It follows that both players' beliefs about each other are condensed in their guesses about the other player's behavior ($\bar{T}_{guess}$ and L_{guess}).

This means that the percentage extent $\bar{T}_{guess}$ to which the advisor considers the investor to behave mistrustingly represents the *advisor's first-order belief* about the investor's mistrust. To the advisor's knowledge, there is always a chance that the investor could follow their advice a or at least use it as a reference point. Hence, there is no reason why the advisor would lie by giving 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 $C_L(a)$ while also potentially reducing their monetary payoff $\pi_A(i)$. For this reason, the advisor's set of best responses $S_{A;BR}$ consists only of strategies that include either giving truthful advice a or lying by overstating the optimal investment i^* to get the investor to overinvest and, therefore, increase their own payoff $\pi_A(i)$ ($a \geq i^* \leftrightarrow L \geq 0$).[25] Anticipating this, a rational investor's guess i^*_{guess} on the optimal investment i^* would be below or equal to the received advice a. This implies that he or

---

[23] 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 will empirically analyze the distribution of pursued strategies. For now, I only suppose that the prior probability distributions of players' preferences for honesty and trust are common knowledge among the players, which enables them to pursue their equilibrium strategies.

[24] Note that due to the continuous message space of my experiment, my design can address the issue of sophisticated deception through truth-telling [39].

[25] 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: $\forall a \in S_{A;BR}: a \in [i^*, i_{max}]$. Note that a completely dishonest advisor gives advice a equal to the maximal investment i_{max} ($a = i_{max} \rightarrow$ if $i^* < i_{max}$: $L > 0$), whereas a completely honest one gives advice a equal to the optimal investment i^* ($a = i^* \leftrightarrow L = 0$).

she would expect the advice a to be either truthful or an overstating lie ($i^*_{guess} \leq a \leftrightarrow L_{guess} \geq 0$). On this basis, the investor would always make an investment i equal or superior to their guessed optimal investment i^*_{guess} ($i \geq i^*_{guess}$). Here the percentage extent L_{guess} to which he or she suspects the advisor to have lied represents the *investor's first-order belief* about the advisor's lying behavior. Expecting nothing but truthful advice a or an overstating lie ($L_{guess} \geq 0$), the investor's set of best responses $S_{I;BR}$ consists only of strategies that involve either following the advice a or mistrusting it by making a risk-reducing investment i below it ($i \leq a \leftrightarrow \overline{T} \leq 0$).[26] As this in turn is known to the advisor, he or she can form their belief about the investor's mistrust respectively ($i_{guess} \leq a \leftrightarrow \overline{T}_{guess} \leq 0$). Please note that this belief $\overline{T}_{guess}$ depends on the given advice a.[27]

With that, the advisor can estimate their utility $U_A(a, i)$ by consulting their beliefs about the investor's type (expressed in $\overline{T}_{guess}$) and making a guess i_{guess} on the investment i. It yields the *advisor's expected utility* $U_A^e\left(a, \overline{T}_{guess}(a)\right)$ as the following function of their advice a and their first-order belief about the investor's mistrusting behavior $\overline{T}_{guess}$:

$$\begin{aligned} U_A^e\left(a, \overline{T}_{guess}(a)\right) &= \pi_A\left(i = i_{guess} = \left(\overline{T}_{guess}(a) + 1\right) * a\right) - C_L(a) \\ &= m_A * \left(\overline{T}_{guess}(a) + 1\right) * a - C_L(a) \end{aligned} \tag{18}$$

with $m_A > 0$ and $\overline{T}_{guess} \leq 0$.[28]

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

$$\max_a \left(U_A^e\left(a, \overline{T}_{guess}(a)\right) \middle| \ a \in [i^*, i_{max}] \right) \quad \rightarrow \quad \frac{\partial\left(C_L(a)\right)}{\partial a} = m_A * \frac{\partial\left(\left(\overline{T}_{guess}(a) + 1\right) * a\right)}{\partial a} \tag{19}$$

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

In order to better understand this maximization problem, it should be noted that the term on the left ($\frac{\partial(C_L(a))}{\partial a}$) solely depends on the advisor's type, whereas the one on the right ($m_A * \frac{\partial\left(\left(\overline{T}_{guess}(a)+1\right)*a\right)}{\partial a}$) depends on their payoff structure as well as their beliefs about the investor's type. This implies that the advisor aims to give advice a so that their preference for honesty (i.e., their sensitivity to change of their moral costs of lying) is balanced in a certain 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).

Solving this maximization problem, finally leads to the *advisor's set of best responses* $S_{A;BR}$:

---

[26] In fact, the investor's set of best responses $S_{I;BR}$ contains only investments i between their guessed optimal investment i^*_{guess} (which is equal to: $\frac{a}{L_{guess}+1}$ by definition) and the advice a. Hence: $\forall i \in S_{I;BR}$: $i \in \left[\frac{a}{L_{guess}+1}, a\right]$.

[27] It should be noted that the advisor's guess i_{guess} on the investment i also depends on the advice a. In particular, since the investor could use the advice a as a reference point for the investment i, the guessed investment i_{guess} should increase monotonously with the given advice a and become zero, if the advice a is zero.

[28] For a more detailed derivation of the advisor's expected utility $U_A^e(a, \overline{T}_{guess}(a))$ and their set of best responses $S_{A;BR}$, see appendix B.1.

$$S_{A;BR} = \left\{ a^s \,\middle|\, a^s, a^{s\prime} \in \left\{ a \in [i^*, i_{max}] \,\middle|\, \frac{\partial\big(C_L(a)\big)}{\partial a} = m_A * \frac{\partial\Big(\big(\bar{T}_{guess}(a)+1\big)*a\Big)}{\partial a} \vee a = i^* \vee a = i_{max} \right\} \wedge \forall a^{s\prime}: U_A^e(a^s) \geq U_A^e(a^{s\prime}) \right\} \tag{20}$$

with $m_A > 0$ and $\bar{T}_{guess} \leq 0$.[29]

Translated to the aforementioned taxonomy of lies and truth-telling, this means that the best response of the *homo moralis* type of advisor would be to engage in either selfish lying, profitable white lying, cooperative truth-telling, or unprofitable truth-telling – depending on their own preference for honesty and their beliefs about the investor's mistrust.

From receiving the advice a the investor learns the upper limit of the optimal investment i^*.[30] As the investor has no further information on the location of the optimal investment i^*, he or she has to consult their beliefs about the advisor's dishonesty (expressed in L_{guess}) by making a guess i^*_{guess} on the value of the true optimal investment i^*. As a result, the *investor's expected utility* $U_I^e\big(a, i, L_{guess}\big)$ can be expressed as the following function of the advice a, the investment i, and the investor's first-order belief about the advisor's lying behavior L_{guess}:

$$\begin{aligned} U_I^e\big(a, i, L_{guess}\big) &= U_I\left(a, i, i^* = i^*_{guess} = \frac{a}{L_{guess}+1}\right) \\ &= \begin{cases} i < \dfrac{a}{L_{guess}+1}: & i_{max} + m_I * i - C_{\bar{T}}(a,i) \\ i \geq \dfrac{a}{L_{guess}+1}: & i_{max} + m_I * \left(2 * \dfrac{a}{L_{guess}+1} - i\right) - C_{\bar{T}}(a,i) \end{cases} \end{aligned} \tag{21}$$

with $m_I > 0$ and $L_{guess} \geq 0$.[31]

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 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 $U_I^e\big(a, i, L_{guess}\big)$ with regard to their investment $i \in \big[i^*_{guess}, a\big]$.[32] It follows:

$$\max_i\left(U_I^e\big(a, i, L_{guess}\big) \,\middle|\, i \in \left[\frac{a}{L_{guess}+1}, a\right]\right) \quad \rightarrow \quad \frac{\partial\big(C_{\bar{T}}(a,i)\big)}{\partial i} = -m_I \tag{22}$$

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

Regarding this maximization problem, it should be noted that the left term ($\frac{\partial(C_{\bar{T}}(a,i))}{\partial i}$) depends on the investor's type, whereas the right one ($-m_I$) is determined by their payoff structure. It follows that the investor seeks to make an investment i so that their 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., the rate at which their payoff decreases when their investment deviates from its optimum).

Solving this maximization problem, yields the *investor's set of best responses* $S_{I;BR}$:

[29] For a more detailed derivation of the advisor's set of best responses $S_{A;BR}$, see appendix B.1.

[30] 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.

[31] For a more detailed derivation of the investor's expected utility $U_I^e\big(a, i, L_{guess}\big)$ and their set of best responses $S_{I;BR}$, see appendix B.2.

[32] Note that: $i^*_{guess} = \frac{a}{L_{guess}+1}$.

$$S_{I;BR} = \left\{ i^s \,\middle|\, i^s, i^{s\prime} \in \left\{ i \in \left[\frac{a}{L_{guess}+1}, a\right] \,\middle|\, \frac{\partial\left(C_{\bar{T}}(a,i)\right)}{\partial i} = -m_I \vee i = \frac{a}{L_{guess}+1} \vee i = a \right\} \wedge \forall i^{s\prime}: U_I^e(i^s) \geq U_I^e(i^{s\prime}) \right\} \tag{23}$$

with $m_I > 0$ and $L_{guess} \geq 0$.[33]

With respect to the earlier introduced taxonomy of mistrust and trust, this implies that the best response of the *homo moralis* type of investor would be to engage in either profitable mistrust, cooperative trust, or benevolent trust – depending on their own preference for trust and their beliefs about the advisor's honesty.

Finally, 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: $L = L_{guess} = \frac{a}{i^*} - 1$ and $\bar{T} = \bar{T}_{guess} = \frac{i}{a} - 1$). Hence, 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 Continuous Deception Game is:

$$S_{equ} = \left\{ (a^s, i^s) \,\middle|\, \left(a^s, a^{s\prime} \in \left\{ a \in [i^*, i_{max}] \,\middle|\, \frac{\partial\left(C_L(a)\right)}{\partial a} = m_A * \frac{\partial\left(i^s(a)\right)}{\partial a} \vee a = i^* \vee a = i_{max} \right\} \wedge \forall a^{s\prime}: U_A^e(a^s) \geq U_A^e(a^{s\prime}) \right) \wedge \left(i^s, i^{s\prime} \in \left\{ i \in [i^*, a^s] \,\middle|\, \frac{\partial\left(C_{\bar{T}}(a^s,i)\right)}{\partial i} = -m_I \vee i = i^* \vee i = a^s \right\} \wedge \forall i^{s\prime}: U_I^e(i^s) \geq U_I^e(i^{s\prime}) \right) \right\} \tag{24}$$

with $m_A > 0$ and $m_I > 0$.[34]

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 3.

Table 3. Possible equilibrium strategy combinations.

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

[33] For a more detailed derivation of the investor's set of best responses $S_{I;BR}$, see appendix B.2.

[34] For a more detailed derivation of this set of game theoretical equilibria S_{equ}, see appendices B.1. to B.3.

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 prefer 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.

3.5.3. Further theoretical implications

After analyzing what rational players would do in the Continuous Deception Game, 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 beliefs about the other player would influence their behavior.

So far, the cost functions ($C_L(a)$ and $C_{\bar{T}}(a,i)$) were defined very generally. More specifically, I argue in favor of diminishing marginal costs of lying and mistrust, respectively. 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. As argued by Engelmann and Fehr [51], 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. Based on these assumptions, 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.[35] With that, the set of game theoretical equilibria S_{equ} can be used to draw conclusions on the impact that each player's first-order belief about the other player has on their own behavior.

On the one hand, the *advisor*'s beliefs about the investor influences their honesty as follows: The more (risk-reducing) mistrusting behavior the advisor expects from their investor ($|\bar{T}_{guess}| \uparrow$ with $\bar{T}_{guess} \leq 0$), the more he or she lies by overstating ($L \uparrow$ with $L \geq 0$).[36] This is because a higher absolute value of the percentage extent of expected mistrust ($|\bar{T}_{guess}| \uparrow$) raises the moral costs of lying that the advisor is willing to accept in equilibrium ($C_L(a)$ accepted by advisor $\uparrow$). Thus, higher expectations of being mistrusted (i.e., a larger extent of expected mistrust $|\bar{T}_{guess}|$) make a rational advisor engage in more dishonest behavior (i.e., a larger extent of lying L).

On the other hand, the *investor*'s beliefs about the advisor influences the extent to which he or she considers investing conservatively in equilibrium. Strictly speaking, this means that the higher the extent to which 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 considers in the first place.[37] In other words, a stronger suspicion of being lied to (expressed in L_{guess}) makes a rational investor consider more mistrusting strategies (i.e., considering a higher possible extent of (risk-reducing) mistrust $|\bar{T}|$ with $\bar{T} \leq 0$). Moreover, the costs of mistrust that the investor is willing to accept in equilibrium ($C_{\bar{T}}(a,i)$ accepted by investor $\uparrow$) increase with the monetary incentive to invest optimally ($m_I \uparrow$). Thus, the investor's decision about their

---

[35] It yields the following conditions:

$$\frac{\partial}{\partial a}\left(\frac{\partial(C_L(a))}{\partial a}\right) = \begin{cases} a < i^*: & \leq 0 \\ a > i^*: & \leq 0 \end{cases}$$ for the advisor's cost function $C_L(a)$ and

$$\frac{\partial}{\partial i}\left(\frac{\partial(C_{\bar{T}}(a,i))}{\partial i}\right) = \begin{cases} i < a: & \leq 0 \\ i > a: & \leq 0 \end{cases}$$ for the investor's cost function $C_{\bar{T}}(a,i)$.

[36] For proof, see appendix B.4.

[37] For proof, see appendix B.4.

investment i ultimately depends on the relation between their preference for trust and their monetary incentive to invest optimally.

As this analysis has demonstrated, the Continuous Deception Game is designed to examine the relationship between honesty and trust in a two-player relationship while taking both players' beliefs about each other into account. In this regard, game theory suggests four classes of equilibrium strategy combinations that consist solely of strategies that are expected to be pursued by rational players. This can serve as a reference for the question of which feasible strategies are more likely to be pursued in this game than others. Finally, it has been shown how such rational players would adapt their behavior to their beliefs about the other player.

4. Design

I conducted ten consecutive rounds of the Continuous Deception Game. This was done in order to analyze the temporal consistency of players' honesty, trust, and their beliefs about each other as well as the relationship between these traits over time. 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.[38] Secondly, I introduced coins as an in-game currency. One hundred of these coins translated to 8 EUR. Thirdly, based on that, I defined the overall maximal investment (i_{max}) to be equal to 100 coins. Finally, I determined the values of the optimal investments (i^*) for all ten rounds. For the first five rounds, this was done randomly. However, to simplify the amounts of the optimal investments, I only allowed optimal investments that were divisible by 5. In the last five rounds, I reused the optimal investment values of the previous five rounds in the exact same order as before. The participants were not informed about this repetition. They were only told that each round's optimal investment was determined randomly in advance and, therefore, its value would most likely change between rounds. The ten optimal investments, which I used in my experiment, were: 50, 70, 25, 35, 10, 50, 70, 25, 35, and 10 coins in this order. Each of these values 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.[39] 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.

[38] 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 [6], since in most treatments of his sender-receiver game he implemented this same profit-loss-ratio.

[39] Note that, as shown, for example, by Mörtenhuber et al. [52], revealing whether prior information is true or not can induce (unwanted) behavioral changes. Therefore, players in my experiment received no feedback until after the last round of the experiment.

5. Experimental procedures

My experiment was conducted on February 1, 2018 at the University of Kassel with a total of 65 participants.[40] 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. The participation in my experiment was voluntary in addition to a lecture on game theory. Since this paper aims to investigate the lying and trust of people in an economic context, I find it an advantage that most subjects of my sample have an economic background and were therefore familiar with the principles of economics. This is reflected in the data obtained in my post-experimental questionnaire, according to which 58.1% of my participants were students in economics, 17.7% in engineering, 9.7% in cultural studies, and 14.5% came from various other fields of study.

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 four stages: (1) the instruction stage, (2) the Continuous Deception Games, (3) the post-experimental questionnaire, and (4) the payoff stage. 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.[41] The participants were allowed to ask questions privately. In (2) the game stage, I conducted ten rounds of the Continuous Deception Game as described in the design section.[42] Afterwards, each participant had to fill out (3) my post-experimental questionnaire. This was also done by using *classEx*. The completion of this questionnaire was a necessary condition for receiving (4) the full payoff at the end of the experiment.[43] The payoffs ranged from 0.8 to 10 EUR with an average of 5.63 EUR. The entire experimental procedure took about one hour. However, it should be noted that the game stage only took about 15 minutes, since each round took less than 90 seconds to complete.

6. Hypotheses derivation

Even though this paper is explorative in large parts, in this section I will formulate my expectations towards both players' behavior and beliefs.

6.1. Overall behavior and beliefs

This subsection does not cover temporal effects yet, which means that all ten rounds of the Continuous Deception Game are considered jointly.[44]

Based on my game theoretical reflections about both players' behavior and first-order beliefs, one would expect the advisors either to lie by overstating the value of the optimal investment in order to make the investors increase the amounts of investments or to tell the truth (i.e., cooperative truth-telling,

---

[40] 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, a sample size of 14 subjects *per group* would already achieve a statistical power of 0.80 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 of suspected lying at the 5%-threshold is 0.99.

[41] For the instructions that I presented to the participants in *classEx* before the game stage started, see appendices C.1.1. (*advisor*) and C.2.1. (*investor*).

[42] For the *classEx* input screens that I used for both players in each round of the experiment, see appendices C.1.2. (*advisor*) and C.2.2. (*investor*).

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

[44] Therefore, aggregated values in this subsection refer to aggregates over all respective participants *and* all ten rounds.

optimal profitable white lying, or selfish lying). In addition, one would suppose that the investors will either engage in risk-reducing mistrust in order to make better investments or to trust their advisors (i.e., cooperative trust, benevolent trust, or profitable mistrust). Therefore, one would assume that both players could anticipate the general orientation of the other player's behavior. While this provides a good starting point for predicting what both players might do, I refrain from formulating hypotheses about these general assumptions, since I will delve into the strategies that I expect both players to pursue in more detail later in this section.

To begin with, I will address the question of how accurate both players' beliefs about the other player's behavior will be. In Peeters et al.'s [41] sender-receiver game, the fraction of senders who believe that the receivers will trust them is close to the fraction of receivers who actually do so (51.16% vs. 58.33%). Since this comparison is not the focus of Peeters et al.'s [41] paper, they do not test this difference for significance. Moreover, in a pre-test of Gneezy's [6] sender-receiver game, he finds that 82% of senders expect the receivers to follow their messages, which is close to the fraction of receivers who actually do so (78%). On this basis, Gneezy [6] assumes that senders correctly anticipate the responses of the receivers. Erat and Gneezy [7] argue that this assumption becomes more plausible with a richer message space, which is certainly given in my Continuous Deception Game. Following this line of argumentation, I expect the *advisors* in my experiment to predict their investors' extent of mistrust mostly accurately, which leads to my first hypothesis:

Hypothesis 1.1 (H1.1). *The advisors' average percentage extent of expected mistrust ($\bar{T}_{guess}$) does not differ significantly from the investors' average percentage extent of mistrust ($\bar{T}$).*

By design, the *investors* in my experiment are provided with less information than the advisors. Even though they can rule out sophisticated deception through truth-telling [39] due to the continuity of the message space in my game,[45] it might be rather difficult for them to correctly estimate their advisors' extent of lying. Peeters et al. [41] observe that the fraction of senders who lie in their sender-receiver game is considerably higher than the fraction of receivers who expect them to do so (72.40% vs. 50.62%). On this basis, I predict that the investors in my experiment will underestimate the extent to which their advisors lie by overstating. Hence:

Hypothesis 1.2 (H1.2). *The advisors' average percentage extent of lying (L) is higher than the investors' average percentage extent of suspected lying (L_{guess}).*

As explained before, in my game, one's behavior and expectations towards the other player determine one's beliefs about the quality of the investment. Therefore, if my expectations about the accuracy of *both players*' beliefs about each other are correct, the following also holds true: Firstly, the advisors' expectations towards the quality of the investment should be relatively accurate. This implies that their estimates of the extent of misinvestment would be close to the actual extent of misinvestment. Secondly, the investors should underestimate the extent to which their investments deviate from the true optimal investment. For these reasons, I predict that the advisors will expect a higher extent of misinvestment than the investors. This leads to my next hypothesis:

Hypothesis 1.3 (H1.3). *The advisors' average percentage extent of expected misinvestment ($F_{A;guess}$) is higher than the investors' average percentage extent of expected misinvestment ($F_{I;guess}$).*

6.2. Sensitivity to in-game variables

In the Continuous Deception Game there are two in-game variables that I expect to be particularly relevant for both players' strategies. In the first place, there is the predefined value of the optimal investment, which I predict to influence the *advisors*' honesty. If the optimal investment is zero, the expected payoff for giving truthful advice is zero, too. Also, the payoff that is expected from truth-telling increases with the given optimal investment. I argue that the advisors would consider it unsatisfying to

[45] Note that 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 [39].

receive a relatively low (or even no) payoff for telling the truth, whereas they would be more satisfied if truth-telling resulted in a relatively high (or even the maximal) payoff. In addition, the advisors could have a minimum level of aspiration for their expected payoff, which they would not want to undercut.[46] For these reasons, I predict that the lower the true optimal investment, the more willing will the advisors be to deviate from the truth and, thus, the higher will be the extent to which they lie by overstating. It follows:

Hypothesis 2.1 (H2.1). *The value of the optimal investment (i^*) is negatively correlated with the average percentage extent of lying (L) per round.*

In the second place, I expect the advice to influence the *investors*' behavior and expectations towards the advisors' honesty. Bear in mind that it is common knowledge that the advisors have a monetary incentive to give advice that overstates the value of the optimal investment. Thus, it makes sense that the investors should associate higher investment advice with more dishonest behavior, which would also be in line with game theoretical predictions. In addition, the investors could use the expected value of the optimal investment (which is exactly half of the maximal possible investment) as a reference point to estimate its true location. Therefore, they could perceive advice that largely differs from this reference point as less likely to be honest (even though it is known that every value of the optimal investment comes up with the same probability). However, the investors should only use this heuristic for advice above this reference point, since they have no reason to believe that their advisors would have understated the value of the optimal investment. For these reasons, I expect that, the higher the received advice, the higher is the extent of lying that the investors suspect. Hence:

Hypothesis 2.2 (H2.2). *The received advice (a) is positively correlated with the average percentage extent of suspected lying (L_{guess}) per advice.*

Furthermore, it is reasonable that the investors would adapt their behavior to these expectations and, therefore, react to a larger extent of suspected lying with more mistrusting behavior. Thus, I predict the investors to engage in more risk-reducing mistrust, the higher their received advice. This leads to my next hypothesis:

Hypothesis 2.3 (H2.3). *The received advice (a) is positively correlated with the average percentage extent of risk-reducing mistrust ($\bar{T}$ with inverted sign) per advice.*

Finally, I expect the *advisors* to anticipate this. For this reason, I hypothesis:

Hypothesis 2.4 (H2.4). *The given advice (a) is positively correlated with the average percentage extent of expected risk-reducing mistrust ($\bar{T}_{guess}$ with inverted sign) per advice.*

6.3. Temporal consistency

Since in my experiment the Continuous Deception Game is repeated over ten consecutive rounds, both players might change their strategies over time. At this point, it should be reminded that advisors and investors were rotated in such a way that no one was matched with the same player twice. This was done in order to prevent common learning effects and backward induction between rounds. With that, from a game theoretical perspective, both players should not change their strategy over time, since there is no reason why their types or their beliefs would change during the experiment. However, from a behavioral perspective, players' preferences and beliefs might not be completely constant over time, which could lead to behavioral changes between rounds. Also, uncertainty about the other players might urge them to reconsider their strategy between rounds. Concerning this matter, Peeters et al. [40] repeated a binary choice-based sender-receiver game over 100 rounds with a procedure to randomly rematch senders and receivers in each round. They find that both the senders' average truth-telling rate and the

[46] The advisors' minimum level of aspiration would be located somewhere between their minimal and their maximal possible payoff.

receivers' average trust rate fluctuate around certain levels over time. This is supported by Sánchez-Pagés and Vorsatz [37] who observe a similar pattern in another sender-receiver game with binary choices, which was repeated 50 times. Based on this, I expect that in my experiment the advisors' extent of lying and the investors' extent of mistrust will be at least partially consistent over time, too. Hence, I predict:

Hypothesis 3.1 (H3.1). *The percentage extent of lying (L) is positively correlated with its time-lagged values.*

Hypothesis 3.2 (H3.2). *The percentage extent of mistrust ($\bar{T}$) is positively correlated with its time-lagged values.*

In addition, I suppose that both players will adequately anticipate this development of the other player's behavior over time. It follows that my next pair of hypotheses is:

Hypothesis 3.3 (H3.3). *The percentage extent of expected mistrust ($\bar{T}_{guess}$) is positively correlated with its time-lagged values.*

Hypothesis 3.4 (H3.4). *The percentage extent of suspected lying (L_{guess}) is positively correlated with its time-lagged values.*

6.4. Strategic orientation based on the relationship between behavior and beliefs

Feasible strategies in the Continuous Deception Game always consist of combinations of behavior and first-order beliefs. On this matter, game theory assumes that one's beliefs can affect one's behavior rather than the other way around. From a behavioral perspective, however, both effective directions appear plausible. On the one hand, strategic *advisors* could adapt their lying behavior to their beliefs about the type of investor they are dealing with. On the other hand, the advisors could adjust their beliefs about their investors to their lying behavior in retrospect in order to avoid cognitive dissonance [53] and maintain a positive self-concept [14]. This could help them to appear honest without being so [54,55], since on paper it would appear as if they would not want to benefit from advising to make an excessive investment but rather prevent a mistrusting investor from overinvesting.[47] This form of cheating rationalization is underlying a much more self-centered or rather self-righteous type of advisor who decides on their behavior primarily based on their individual preferences and then adjusts their beliefs accordingly. What both of these perspectives have in common is that they predict the advisors' extent of expected mistrust to correlate with their extent of lying. In line with this, López-Pérez and Spiegelman [12] find in their sender-receiver game with binary choices that the beliefs of the senders are correlated with their decision to lie. Moreover, Peeters et al. [41] provide evidence for belief-based truth-telling in their sender-receiver game, too. For these reasons, I expect the following:

Hypothesis 4.1 (H4.1). *The percentage extent of expected risk-reducing mistrust ($\bar{T}_{guess}$ with inverted sign) is positively correlated with the percentage extent of lying (L).*

In addition, I anticipate the advisors to engage in more strategic than self-righteous decision making. This implies that their first-order beliefs about the investors would affect their behavior rather than the other way around. As a result of the fast sequence of rounds in my experiment, the short-term relationship between the advisors' behavior and beliefs might be visible between rounds over time. Thus, I expect that the advisors' expectations of being mistrusted by their investors in previous rounds (i.e.,

[47] In this way, the advisors could convince themselves that their lies are not hurting their investors (for a comparable type of self-deception in a sender-receiver game, see Gneezy et al. [45]). In my experiment, this could apply in particular to profitable white lies (i.e., lies where the advisors overstate the value of the optimal investment and at the same time expect their investors to (at least partially) compensate the lie by engaging in risk-reducing mistrust). If in these cases the advisors adapted their beliefs about the investors' mistrust to their own lying behavior in retrospect, this type of lying could not be considered as entirely "white" anymore.

their time-lagged extent of expected mistrust) will predict their lying behavior in present rounds (i.e., their extent of lying) – rather than the other way around. Hence:

Hypothesis 4.2 (H4.2). *The time-lagged correlation between the percentage extent of expected mistrust ($\bar{T}_{guess}$) and the percentage extent of lying (L) is stronger when the first-mentioned is preceding the last-mentioned ($\bar{T}_{guess} \rightarrow L$) rather than the other way around ($L \rightarrow \bar{T}_{guess}$).*

If this assumption is correct, this would speak in favor of strategic behavior, which is in line with rational decision making as it is supposed in the game theoretical approach. In this sense, rational advisors would engage neither in strategies that are against their internal preferences nor inconsistent with their beliefs about the investors. As a result, such advisors would always engage in equilibrium strategies. Therefore, I assume that advisors who engage in strategies that could potentially be equilibrium strategies are rational. Based on my expectations of strategic decision making, I predict the advisors to pursue rational strategies more often than other strategies. It follows:

Hypothesis 4.3 (H4.3). *The advisors pursue rational strategies that are potential equilibrium 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 $\bar{T}_{guess} > 0$).*

It is reasonable to suppose that the *investors* would adapt their behavior and beliefs about their advisors to one another, too. From a game theoretical perspective, they should adjust their behavior to the extent of lying that they expect from their advisors. More precisely, the higher the extent of suspected lying, the more mistrusting investments the investors would potentially consider in equilibrium. In addition, trusting beliefs, i.e., expectations about another person's honesty, are closely related to the realization of trusting behavior [30]. This is in line with Peeters et al. [41] who find in their sender-receiver game with binary choices that the probability that the receivers choose trust increases with the probability that they believe the senders to tell the truth. On this basis, I expect the investors' extent of suspected lying to be correlated with their extent of mistrust:

Hypothesis 4.4 (H4.4). *The percentage extent of suspected lying (L_{guess}) is positively correlated with the percentage extent of risk-reducing mistrust ($\bar{T}$ with inverted sign).*

Analogous to the advisors, I argue that some investors will try to engage in strategic decision making, i.e., that they will base their investments on their beliefs about the advisors. However, by design, the investors have less information and are faced with more uncertainty than the advisors. This could also lead to more intuitive decision making. Therefore, some investors could make less strategic investments and then adapt their beliefs about their advisors to their behavior in retrospect. Even though I expect this to happen to some extent, I believe that the investors will engage in more strategic than intuitive decision making. Again, this relationship between the investors' behavior and their first-order beliefs might be visible between rounds over time. That would imply that the investors' behavior in one round (i.e., their extent of mistrust) might be based on their expectations about their advisors' dishonesty in previous rounds (i.e., the time-lagged extent of suspected lying). It yields the following hypothesis:

Hypothesis 4.5 (H4.5). *The time-lagged correlation between the percentage extent of suspected lying (L_{guess}) and the percentage extent of mistrust ($\bar{T}$) is stronger when the first-mentioned is preceding the last-mentioned ($L_{guess} \rightarrow \bar{T}$) rather than the other way around ($\bar{T} \rightarrow L_{guess}$).*

Furthermore, similar to the advisors, I consider investor strategies that could potentially be equilibrium strategies as rational. Based on that, as I expect the investors to make primarily strategic investments, I predict that they will pursue rational strategies more often than other strategies. From this follows my last hypothesis:

Hypothesis 4.6 (H4.6). *The investors pursue rational strategies that are potential equilibrium strategies (with $F_{I;guess} \geq 0$ and $\bar{T} \leq 0$) disproportionately more often than other strategies (with $F_{I;guess} < 0$ or $\bar{T} > 0$).*

I now turn to an examination of my results in light of these hypotheses.

7. Results

This section contains my findings from the Continuous Deception Game. It is divided into five subsections: (*7.1.*) the main results on players' honesty, trust, and their first-order beliefs, (*7.2.*) an analysis of players' sensitivity to the most important in-game variables, (*7.3.*) the testing of players' behavior and beliefs for temporal consistency, (*7.4.*) an analysis of both players' strategic orientation based on the relationship between their behavior and first-order beliefs, and finally (*7.5.*) the testing of the consistency of my ex ante classification of both players' strategies with their ex post self-assessment of their own behavior.

7.1. Main results on honesty, trust, and beliefs

In this subsection I present overall results concerning the seven key variables of the Continuous Deception Game. Therefore, I begin by examining the advisors' lying behavior and the investors' expectations towards it (*(Suspected) lying*). This is followed by an analysis of the investors' mistrust and the advisors' expectations about it (*(Expected) mistrust*). Finally, I examine the outcome of the investment and both players' expectations towards it (*(Expected) misinvestment*). All results in this subsection are non-temporal. To this end, all ten rounds of the experiment are considered jointly. Hence, all values in this subsection refer to aggregates over all ten rounds.

7.1.1. (Suspected) lying

Table 4 compares the observed lying behavior of the advisors to the expectations of the investors towards it. Firstly, it shows the proportions of different types of (suspected) lies on average over all ten rounds.[48] Secondly, it displays the average percentage extent of (suspected) lying over all ten rounds.

Table 4. Proportion and extent of (suspected) lying.

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.

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 lying do not differ significantly (Mann-Whitney U test: p = 0.960).[49] However, the investors significantly overestimated the proportion of advisors who understated the optimal investment: They

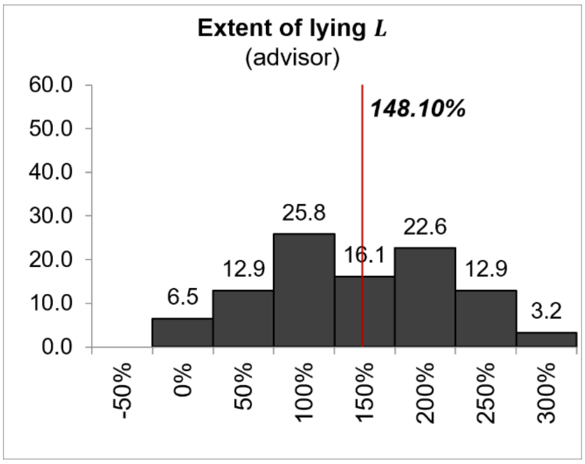
[48] The *proportion of lying* refers to the percentage of advisors who lied (i.e., the relative frequency of observed $L \neq 0$), calculated as an average over all ten rounds. Analogously, the *proportion of suspected lying* is defined as the percentage of investors who suspected their advisors to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

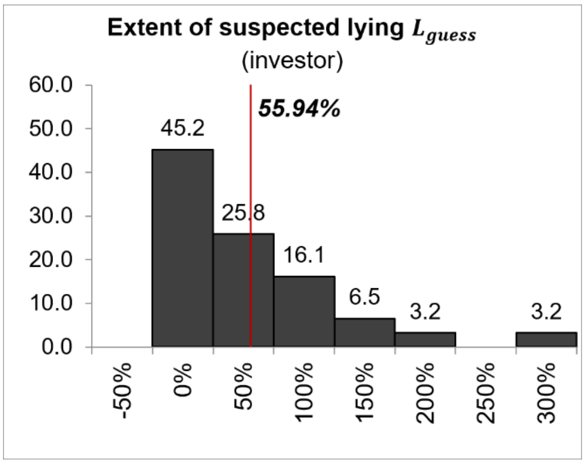
[49] Solely based on this information, one might (falsely) conclude that the investors' beliefs about their advisors' dishonesty were highly accurate. However, taking the extents of lying and suspected lying into account will reveal that this assessment is not correct. This is important since the findings of binary choice-based sender-receiver games are usually based solely on such proportions.

suspected 20.00% of all pieces of advice to be understating lies, whereas only 1.29% of them actually were (Mann-Whitney U test: p < 0.001). In addition, the variance of the rate at which the advisors told understating lies is significantly lower than the variance of the rate at which the investors suspected such lies (non-parametric Levene's test: p < 0.001). This is due to the fact that none of the advisors engaged in lying by understating more than twice, while 35.48% of all investors suspected their advisors to do so. This difference is significant (Fisher exact test: p < 0.001). Moreover, the proportion of overstating lies (77.10%) was significantly higher than the investors suspected (58.39%) (Mann-Whitney U test: p = 0.004). Here the proportion of advisors (77.42%) who overstated the true value of the optimal investment by an average of at least 100% is significantly higher than the proportion of investors (12.90%) who suspected them to do so (Fisher exact test: p < 0.001). It should be noted, however, that the advisors lied to the fullest possible extent in only 8.71% of all cases. I will come back to this later when I discuss the sensitivity of their beliefs about their investors' mistrust to their given advice.

Finding 1.1. The advisors lied in 78.39% of cases and the exact same proportion of investors suspected them to do so. Most (suspected) lies were overstating lies.

Finding 1.2. The investors overestimated the proportion of advisors who lied by understating, while they underestimated the proportion of advisors who lied by overstating. In particular, the investors underestimated the proportion of extremely dishonest advice.





(**a**) (**b**)

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}.

For a more detailed analysis, 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).[50] 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 percentage extent of lying appears to be not too far from a normal distribution (one-sample Kolmogorov-Smirnov test: p = 0.200), 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 a different pattern to estimate their advisors' lying behavior than the advisors actually did. In the second place, the peak of the frequency distribution of the investors'

[50] *Descriptive results for Figure 5*: As can be read from the distribution of the percentage *extent of lying* (L) in Figure 5a, the advisors lied by overstating to an average extent of 148.10%. Here, the percentage extent of lying is widely spread (standard deviation of 73.10%) within the range from 21.95% to 285.43%. Turning to the distribution of the percentage *extent of suspected lying* (L_{guess}), Figure 5b shows that the investors expected their advisors to have lied by overstating to an average extent of 55.94%. Again, the percentage extent of suspected lying is widely spread (standard deviation of 68.99%) and lies within the range from -22.22% to 288.10%.

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 indicates 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). Therefore, I find strong support for *hypothesis H1.2*. It follows that the investors underestimated the percentage extent of lying on average by 92.16% (L vs. L_{guess}).

Finding 1.3. The advisors lied by overstating ($L \gg 0$) to an average extent of 148.10%, while the investors expected them to do so ($L_{guess} > 0$) by 55.94% on average.

Finding 1.4. The investors underestimated the extent to which the advisors lied by overstating ($L \gg 0$ vs. $L_{guess} > 0$).

7.1.2. (Expected) mistrust

Table 5 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.[51] In addition, it reports the average percentage extent of (expected) mistrust over all ten rounds.[52]

Table 5. Proportion and extent of (expected) mistrust.

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

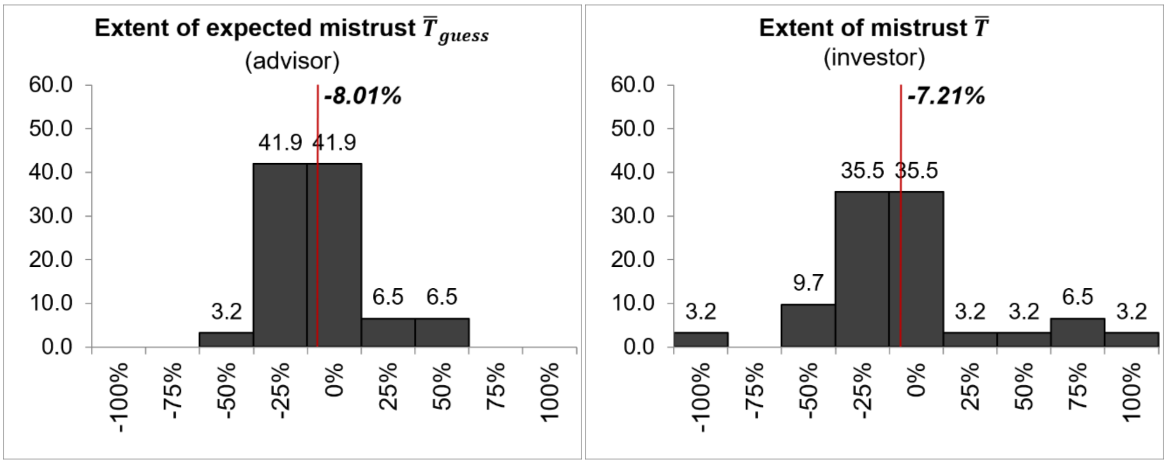
It can be seen that the advisors expected 68.71% of all investments to be mistrusting, while 72.26% of them actually were. This minor difference is not significant (Mann-Whitney U test: p = 0.701). In line with this, the advisors predicted the proportions of different types of mistrust highly accurately: Firstly, the proportion of advisors who expected risk-reducing mistrust is only slightly higher than the proportion of investors who engaged in such mistrust (58.06% vs. 57.10% of all investments; Mann-Whitney U test: p = 1.000). Secondly, the advisors expected 10.65% of all investments to be based on risk-seeking mistrust, while 15.16% of them actually were. Again, this difference is not significant (Mann-Whitney U test: p = 0.273).

Finding 1.5. The investors mistrusted their advisors in 72.26% of cases, while the advisors expected them to do so in 68.71% of cases. In most cases the (expected) mistrust was risk-reducing.

[51] The *proportion of mistrust* is defined as 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. Analogously, 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.

[52] Bear in mind 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.

Finding 1.6. The advisors predicted the proportions of different types of mistrust highly accurately.



(**a**) (**b**)

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 $\bar{T}$.

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).[53] 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. 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-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 to 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). On this basis, *hypothesis H1.1* cannot be rejected. It should be noted, however, that the variance of the percentage extent of expected mistrust is significantly lower than the variance of the percentage extent of mistrust (non-parametric Levene's test: p = 0.048). This suggests that the advisors did not anticipate that some investors would mistrust them by making either extremely risk-reducing or extremely risk-seeking investments. Overall, however, the advisors barely misjudged the extent of their investors' mistrust.

Finding 1.7. The investors mistrusted their advisors ($\bar{T} < 0$) to an average extent of -7.21%, while the advisors expected them to do so ($\bar{T}_{guess} < 0$) by -8.01% on average. Thus, the (expected) mistrust tended to be risk-reducing.

[53] *Descriptive results for Figure 6*: The distribution of the percentage *extent of expected mistrust* ($\bar{T}_{guess}$) in Figure 6a reveals that the advisors expected their investors to mistrust them to an average extent of -8.01%, which implies that on average they expected their investors to tend to engage in risk-reducing mistrust. It should be noted that the percentage extent of expected mistrust is spread within the range from -44.11% to 62.17% (with a standard deviation of 22.60%). This indicates that the advisors expected not only risk-reducing but also risk-seeking mistrust from their investors. Concerning the distribution of the percentage *extent of mistrust* ($\bar{T}$) in Figure 6b, it can be seen that the investors mistrusted their advisors to an average extent of -7.21%. This shows that on average the investors tended to engage in risk-reducing mistrust. However, the percentage extent of mistrust is widely spread (standard deviation of 37.96%) within the range from -87.61% to 96.98%, which again includes both risk-reducing and risk-seeking mistrust.

Finding 1.8. The advisors' expectations towards the extent of their investors' mistrust were exceptionally accurate ($\bar{T}_{guess} < 0$ vs. $\bar{T} < 0$).

7.1.3. (Expected) misinvestment

Table 6 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.[54] Secondly, it shows both players' average percentage extent of expected misinvestment over all ten rounds.

Table 6. Proportion and extent of expected misinvestment.

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

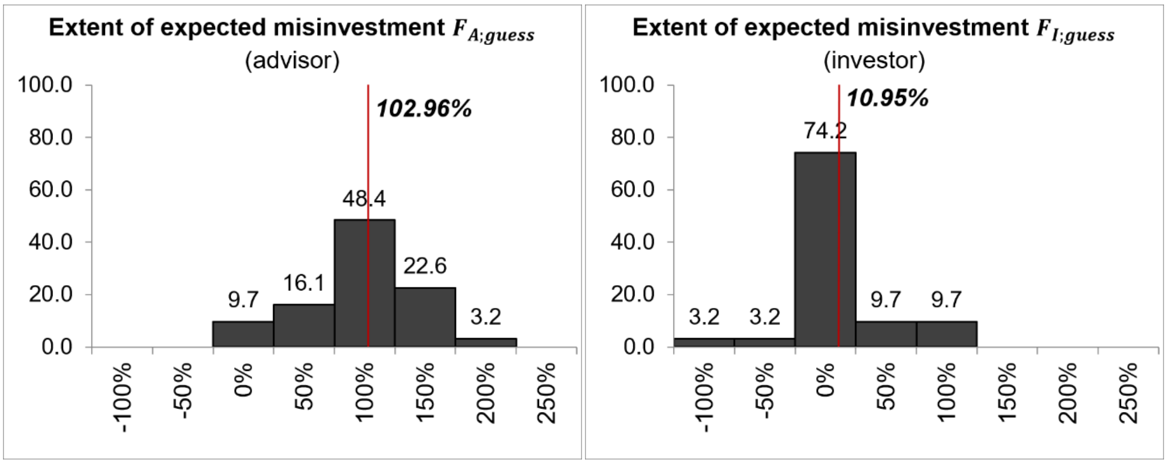
[1] Mann-Whitney U test to compare the mean rank of the respective concerned variables between both players.

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).

Finding 1.9. The advisors expected 73.22% of all investments to be non-optimal (of which most were expected to be overinvestments), while the investors expected 57.42% of their investments to be non-optimal (of which only about half were expected to be overinvestments).

Finding 1.10. The advisors expected fewer under- and more overinvestments than the investors.

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



(**a**) (**b**)

Figure 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 $F_{I;guess}$.

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).[55] Comparing both distributions shows that the expectations of the advisors towards the quality of investments differed greatly from those of the investors: Firstly, both samples apparently do not follow the same distribution (two-sample Kolmogorov-Smirnov test: $p < 0.001$). While the advisors' percentage extent of expected misinvestment seems to be not too far from a normal distribution (a one-sample Kolmogorov-Smirnov test is not significant: $p = 0.200$), the investors' percentage extent of expected misinvestment is not normally distributed (one-sample Kolmogorov-Smirnov test: $p < 0.001$). This difference in both distributions demonstrates that the investors followed a different pattern to estimate the quality of their investments than the advisors. Secondly, the peak of the frequency distribution of the investors' percentage extent of expected misinvestment is close to zero (at around $F_{I;guess} = 0$), whereas the corresponding frequency distribution of the advisors peaks at a much higher level ($F_{A;guess} \gg 0$), which is at around 100%. It follows that optimal investments were a much stronger reference point for the investors than for the advisors. Finally, 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: $p < 0.001$). This finding is consistent with *hypothesis H1.3*. Hence, 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 1.11. The advisors expected their investors to overinvest ($F_{A;guess} \gg 0$) to an average extent of 102.96%, while the investors expected to do so ($F_{I;guess} > 0$) by only 10.95% on average.

Finding 1.12. The advisors expected a higher extent of misinvestment than the investors ($F_{A;guess} \gg 0$ vs. $F_{I;guess} > 0$).

[55] *Descriptive results for Figure 7*: It can be seen in the distribution of the *advisors*' percentage *extent of expected misinvestment* ($F_{A;guess}$) in Figure 7a that the advisors expected their investors to overinvest by 102.96% on average. Here, their percentage extent of expected misinvestment is widely spread (standard deviation of 47.35%) within the range from 17.57% to 204.86%. This means that all advisors expected the average of the investments that were made by their investors to be too high. Turning to the distribution of the *investors*' percentage *extent of expected misinvestment* ($F_{I;guess}$), Figure 7b shows that the investors expected to overinvest by an average extent of 10.95%. Note that their percentage extent of expected misinvestment ranges from -76.09% to 124.48% (with a standard deviation of 41.06%), which includes expectations of making both under- and overinvestments.

Table 7. Proportion and extent of misinvestment.

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%

In order to assess the quality of both players' estimates of the outcomes of the investments, Table 7 provides an overview of the observed quality of investments. Firstly, it shows the proportions of different types of investments on average over all ten rounds.[56] Secondly, 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 was due to the fact that 24.52% of all investments were underinvestments and 59.35% overinvestments.

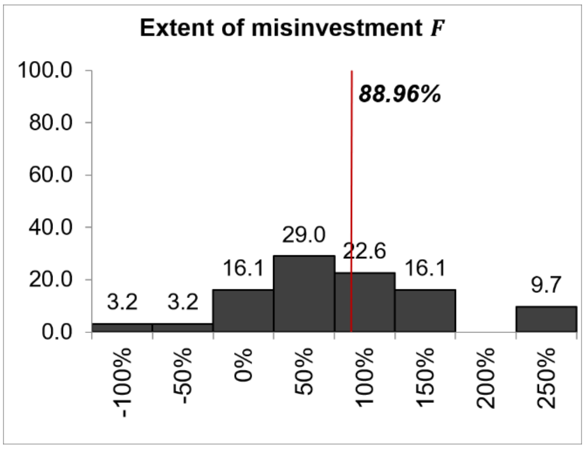


Figure 8. Distribution of the extent of misinvestment F.

For a more detailed analysis, Figure 8 visualizes the distribution of the percentage extent of misinvestment (F).[57] 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 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 1.13. The proportion of non-optimal investments is 83.87%, most of them being overinvestments. As a result, the average percentage extent of misinvestment is 88.96%.

Finding 1.14. The advisors overestimated the extent of misinvestment only moderately, while the investors largely underestimated it.

[56] 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.
[57] *Descriptive results for Figure 8*: The distribution of the percentage *extent of misinvestment* (F) shows that the investors overinvested by an average extent of 88.96%. The underlying distribution is widely spread (standard deviation of 79.09%) and ranges from -83.14% to 273.71%. Moreover, it peaks at a moderately high level, around 50%.

7.1.4. Summary

Table 8 provides an overview of the testing of my main hypotheses about players' honesty, trust, and beliefs. It shows that my general expectations about both players were met.

Table 8. Summary of main hypotheses testing.

Hypothesis	Concerned variables and prediction	Difference of averages (1st-2nd)	p-value (two-sided Mann-Whitney U test[1])	Result
H1.1	$\bar{T}_{guess} \approx \bar{T}$	0.80%	0.767	Not rejected
H1.2	$L > L_{guess}$	92.16%	< 0.001	Supported
H1.3	$F_{A;guess} > F_{I;guess}$	92.01%	< 0.001	Supported

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

Most of the *advisors* lied by overstating (see Finding 1.1), which they did to a relatively high extent (see Finding 1.3). Moreover, most of them expected their investors to mistrust them (see Finding 1.5). In particular, the advisors predicted a slight tendency to risk-reducing mistrust from their investors, which means that on average they expected investments below their given advice (see Finding 1.7). As a result of their high extent of lying and their much lower extent of expected mistrust, the advisors expected their investors to overinvest in most of the cases (see Finding 1.9) and to a high extent (see Finding 1.11). These predictions of the overall quality of investments were decent (see Finding 1.14), since the advisors predicted the proportions of different types of mistrust and the extent of mistrust highly accurately (see Finding 1.6 and 1.8; Hypothesis H1.1).

Most of the *investors* engaged in mistrusting behavior (see Finding 1.5). In general, they had a tendency towards risk-reducing mistrust, meaning that on average they made investments below their received advice (see Finding 1.7). This might be due to the fact that the investors expected most of their advisors to lie by overstating (see Finding 1.1). In line with this, they expected their advisors to overstate the true value of the optimal investment to a moderately high extent (see Finding 1.3). However, the investors still underestimated the proportion and extent of overstating lies (see Finding 1.2 and 1.4; Hypothesis H1.2). As a consequence, most of their investments turned out to be overinvestments (see Finding 1.13). This was not foreseen by the investors. In fact, they expected a moderately high proportion of non-optimal investments (see Finding 1.9) but predicted a low extent of misinvestment (see Finding 1.11). Thus, they misjudged the quality of their investments by underestimating their actual extent of misinvestment (see Finding 1.14).

The investors estimated the dishonesty of their advisors less accurately than the advisors estimated their investors' mistrust. Therefore, the advisors predicted the overall quality of investments much better than the investors. As a result, the advisors expected a higher extent of misinvestment than the investors (see Finding 1.12; Hypothesis H1.3). This is reflected in the fact that the advisors expected fewer under- and more overinvestments than the investors (see Finding 1.10).

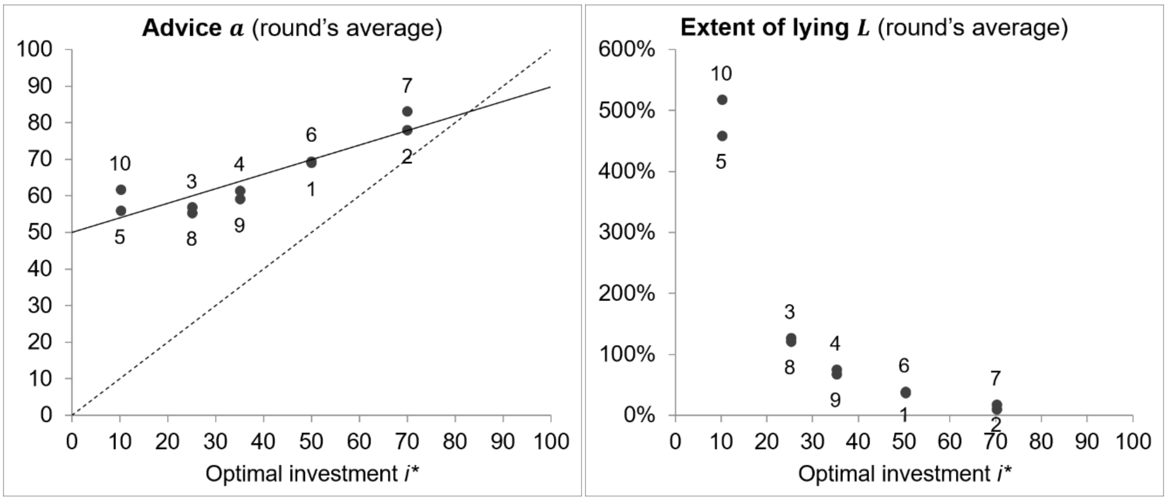
7.2. Sensitivity of honesty, trust, and beliefs to the optimal investment and the advice

In this subsection I intend to test both players' sensitivity to the most important in-game variables, i.e., the predefined optimal investment and the given or received advice. I begin with the advisors (*Lying and expected mistrust*) and then turn to the investors (*Mistrust and suspected lying*).

7.2.1. Lying and expected mistrust (advisors)

In order to analyze the sensitivity of the *advisors' lying behavior* to the predefined optimal investment in more detail, Figure 9 displays two scatter plots. The plot on the left (9a) illustrates the relationship between the average advice (a) per round and the corresponding optimal investment (i^*).

The one on the right (9b) visualizes the relationship between the average percentage extent of lying (L) per round and the corresponding optimal investment (i^*).[58]



(**a**) (**b**)

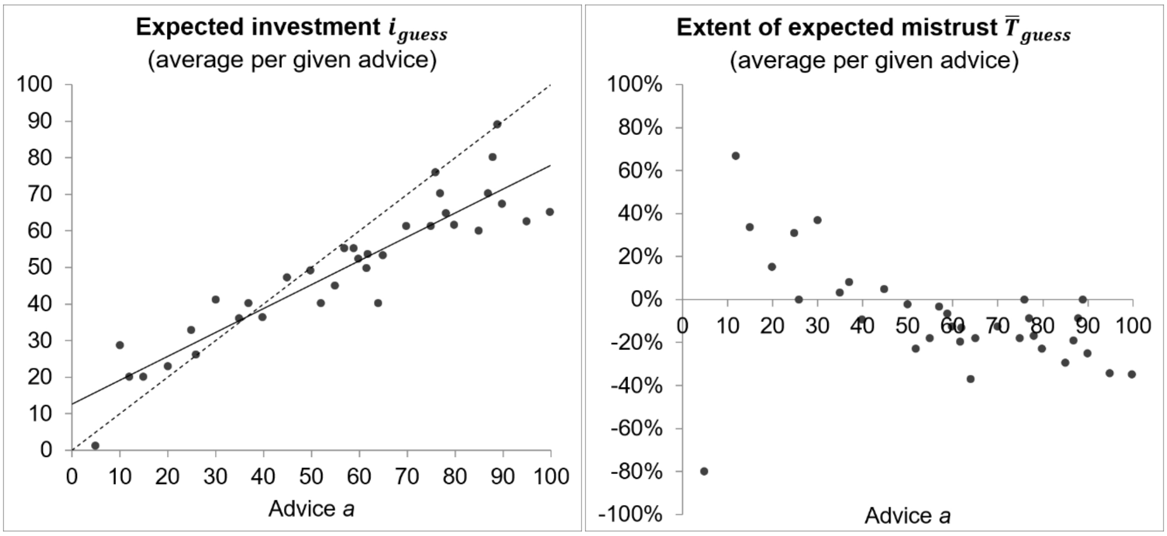
Figure 9. Sensitivity of the advisors' lying behavior to the optimal investment: (**a**) Advice and optimal investment; (**b**) Extent of lying and optimal investment.

As can be read from the scatter plot in Figure 9a, the average advice significantly increases with the value of the optimal investment (Spearman's rank correlation: ρ = 0.812 with p = 0.004). The solid line in the plot illustrates the line of best fit. Notably, this line intersects the ordinate at 49.93, which is extraordinarily close to the midpoint of possible advice ($\frac{i_{max}}{2} = 50$). This suggests that the advisors used this midpoint as a reference point for their advice, indicating that their minimum level of aspiration for their expected payoff corresponded to this midpoint. In order to compare the observed behavior to the truthful one, the dotted line represents the hypothetical line of completely truthful advice (which would imply: $L = 0$).[59] All points in the plot are located above this line, meaning that on average the advisors lied by overstating in each round ($L > 0$). In addition, it can be seen that with lower optimal investments both lines increasingly diverge from each other. Thus, the lower the true optimal investment, the more willing the advisors were to deviate from the truth. This is reflected in the scatter plot in Figure 9b, which demonstrates that the round's average percentage extent of lying significantly decreases with the value of the corresponding optimal investment (Spearman's rank correlation: ρ = -0.985 with p < 0.001). From this, it follows that the lower the optimal investment, the higher the extent to which the advisors lied by overstating its value. Therefore, I find strong support for *hypothesis H2.1*.

Finding 2.1. The lower the optimal investment, the more willing the advisors were to lie by overstating its value and, therefore, the higher was the extent of their overstatement.

Figure 10 consists of two scatter plots, which can be used to examine the sensitivity of the *advisors' first-order beliefs* about their investors' mistrust to their given advice in more detail. The plot on the left (10a) displays the relationship between the average investment that the advisors expected from their investors (i_{guess}) per given advice and the corresponding advice (a). The one on the right (10b) shows the relationship between the average percentage extent of expected mistrust ($\bar{T}_{guess}$) per given advice and the corresponding advice (a).[60]

[58] By design each optimal investment was used twice with a lag of five rounds. I will come back to this in more detail when I analyze my data for temporal consistency.
[59] Note that points below this hypothetical line represent *lying by understating* ($L < 0$), whereas points above this line represent *lying by overstating* ($L > 0$).
[60] For the purpose of illustration, only the most relevant section of the plot in Figure 10b is displayed.



(**a**) (**b**)

Figure 10. Sensitivity of the advisors' first-order beliefs about their investors' mistrust to their given advice: (**a**) Expected investment and advice; (**b**) Extent of expected mistrust and advice.

As can be seen in the scatter plot in Figure 10a, the average expected investment per advice significantly increases with the value of the given advice (Spearman's rank correlation: $\rho = 0.930$ with $p < 0.001$). This comes as no surprise, as it shows that the advisors expected their investors to use their advice as a reference point for the sizes of the investments. The solid line in the plot represents the line of best fit. In comparison, the dotted line represents the hypothetical line on which the advisors would have expected their investors to make completely trusting investments (which would imply: $\bar{T}_{guess} = 0$).[61] Both lines intersect slightly above one-third of the highest possible advice,[62] suggesting that the advisors had different expectations towards their investors' mistrust for lower than for higher advice. For low-level advice (below or equal to the intersection point of both lines) the advisors expected investments below their given advice in only 20.37% of cases, while for high-level advice (above the intersection point) they expected the same in 66.02% of cases. This difference is significant (Fisher exact test: $p < 0.001$). Furthermore, the scatter plot in Figure 10b shows that the average percentage extent of expected mistrust per advice significantly decreases with the value of the given advice (Spearman's rank correlation with outlier-cleaned values: $\rho = -0.728$ with $p < 0.001$). Hence, I find strong support for *hypothesis H2.4*. In line with this, lower advice was primarily associated with expectations of risk-seeking mistrust ($\bar{T}_{guess} > 0$), while higher advice was primarily associated with expectations of risk-reducing mistrust ($\bar{T}_{guess} < 0$).

Finding 2.2. The higher the given advice, the less risk-seeking and the more risk-reducing mistrust the advisors expected from their investors.

Most interestingly, when considering only cases in which the advisors gave very high advice (i.e., advice equal or superior to 85), the correlation between the given advice and the expected investments is close to zero and not significant (Spearman's rank correlation: $\rho = -0.024$ with $p = 0.828$). However, it remains strong and highly significant when considering all cases except those where the advisors gave such very high advice (Spearman's rank correlation: $\rho = 0.669$ with $p < 0.001$). 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' mistrust to compensate particularly high advice above a certain level in such a

[61] Note that points below this hypothetical line represent *expectations of risk-reducing mistrust* ($\bar{T}_{guess} < 0$), whereas points above this line represent *expectations of risk-seeking mistrust* ($\bar{T}_{guess} > 0$).

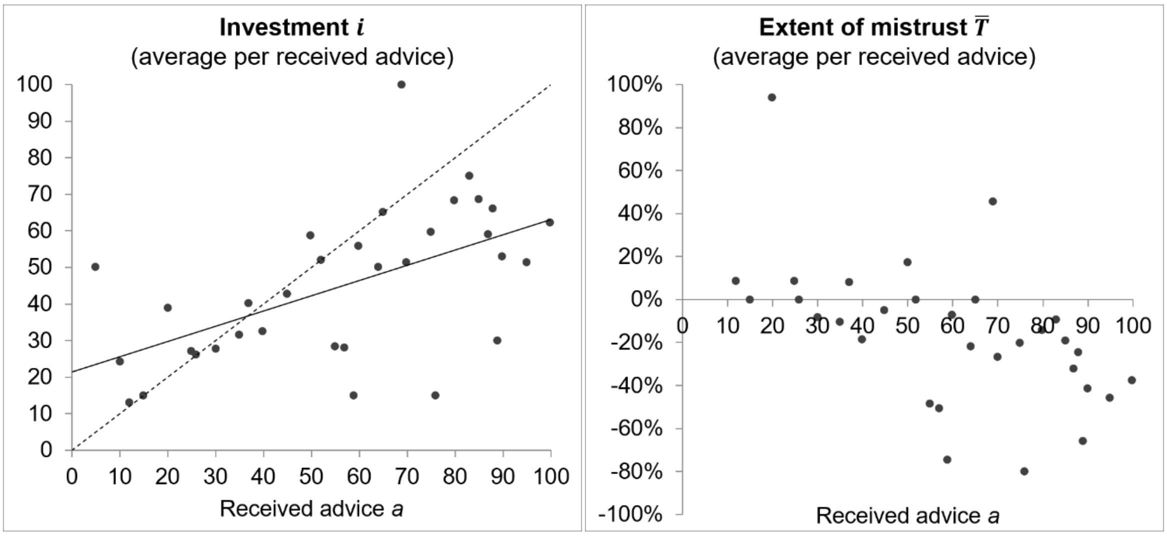
[62] Both lines intersect at 36.43.

way that giving higher advice would not lead to higher investments and not deliver a higher payoff. This may partially explain why the advisors lied to the fullest possible extent in only 8.71% of cases.

Finding 2.3. The advisors expected their investors to mistrust particularly high advice in such a way that giving higher advice above a certain level would not lead to higher investments.

7.2.2. Mistrust and suspected lying (investors)

Figure 11 consists of two scatter plots, which allow for a detailed analysis of the sensitivity of the *investors' mistrusting behavior* to their received advice. The plot on the left (11a) visualizes the relationship between the average investment (i) per received advice and the corresponding advice (a). The one on the right (11b) shows the relationship between the average percentage extent of mistrust ($\bar{T}$) per received advice and the corresponding advice (a).[63]



(**a**) (**b**)

Figure 11. Sensitivity of the investors' mistrusting behavior to their received advice: (**a**) Investment and advice; (**b**) Extent of mistrust and advice.

As can be read from the scatter plot in Figure 11a, the average investment significantly increases with the value of the received advice (Spearman's rank correlation: $\rho = 0.621$ with $p < 0.001$). The solid line represents the line of best fit. In order to compare the observed behavior to the trusting one, the dotted line represents the hypothetical line on which the investors would have exactly followed their received advice (which would imply: $\bar{T} = 0$).[64] Both lines intersect slightly above one-third of the highest possible advice,[65] indicating that the investors' mistrust differed for lower and higher advice. For low-level advice (below or equal to the intersection point of both lines), the investors made investments below their received advice in only 29.51% of cases, while for high-level advice (above the intersection point) they did the same in 63.86% of cases. This difference is significant (Fisher exact test: $p < 0.001$). Moreover, the scatter plot in Figure 11b shows that the average percentage extent of mistrust per advice significantly decreases with the value of the received advice (Spearman's rank correlation with outlier-cleaned values: $\rho = -0.682$ with $p < 0.001$). This finding offers strong support for *hypothesis H2.3*. In line with this, investors who received lower advice tended to engage in risk-seeking mistrust ($\bar{T} > 0$), whereas investors who received higher advice tended to engage in risk-reducing mistrust ($\bar{T} < 0$).

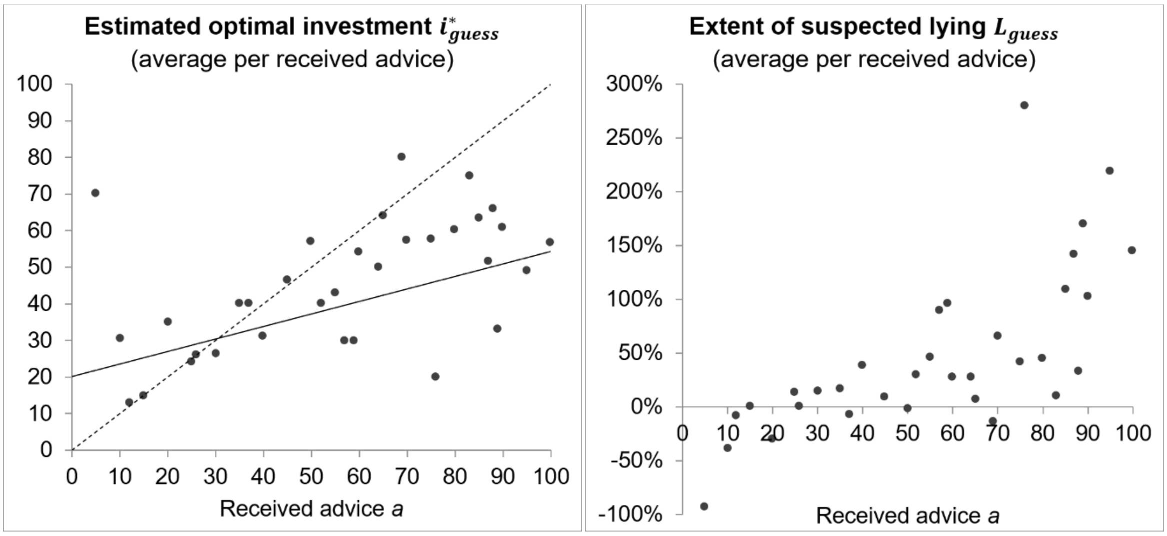
[63] For the purpose of illustration, only the most relevant section of the plot in Figure 11b is displayed.
[64] Note that points below this hypothetical line represent *risk-reducing mistrust* ($\bar{T} < 0$), whereas points above this line represent *risk-seeking mistrust* ($\bar{T} > 0$).
[65] Both lines intersect at 36.80.

Finding 2.4. The higher the received advice, the less the investors engaged in risk-seeking and the more in risk-reducing mistrust.

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 remains strong and highly significant when cases in which the investors received such very high advice are not considered (Spearman's rank correlation: $\rho = 0.549$ with $p < 0.001$). By contrast, the 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 2.5. The investors mistrusted particularly high advice in such a way that they did not make higher investments, when receiving higher advice above a certain level.

For a more detailed analysis of the sensitivity of the *investors' first-order beliefs* about their advisors' lying behavior to their received advice, Figure 12 visualizes two scatter plots. The plot on the left (12a) displays the relationship between the investors' average estimate of the optimal investment (i^*_{guess}) per received advice and the corresponding advice (a). The one on the right (12b) illustrates the relationship between the average percentage extent of suspected lying (L_{guess}) per received advice and the corresponding advice (a).



(**a**) (**b**)

Figure 12. Sensitivity of the investors' first-order beliefs about their advisors' lying behavior to their received advice: (**a**) Estimated optimal investment and advice; (**b**) Extent of suspected lying and advice.

The scatter plot in Figure 12a reveals that the average estimated optimal investment significantly increases with the value of the received advice (Spearman's rank correlation: $\rho = 0.526$ with $p = 0.002$). The solid line represents the line of best fit. This demonstrates that the investors at least partially based their estimates of the optimal investment on their received advice. The dotted line can be used as a reference, since it represents the hypothetical line on which the investors would have expected to receive completely truthful advice (which would imply: $L_{guess} = 0$).[66] As can be seen, both lines intersect slightly below one-third of the highest possible advice,[67] suggesting that the investors had different

[66] Note that points below this hypothetical line represent *expectations of being lied to by overstating* ($L_{guess} > 0$), whereas points above this line represent *expectations of being lied to by understating* ($L_{guess} < 0$).

[67] Both lines intersect at 30.68.

beliefs about their advisors' lying behavior for lower advice as opposed to higher advice. For low-level advice (below or equal to the intersection point of both lines) the investors expected the optimal investment to be lower than their received advice in only 19.44% of cases, whereas for high-level advice (above the intersection point) they expected the same in 63.50% of cases. This difference is significant (Fisher exact test: $p < 0.001$). In line with this, the scatter plot in Figure 12b illustrates that the average percentage extent of suspected lying significantly increases with the value of the received advice (Spearman's rank correlation: $\rho = 0.773$ with $p < 0.001$). Based on this, I find strong support for *hypothesis H2.2*. It can be concluded that investors who received lower advice tended to suspect to be lied to by understating ($L_{guess} < 0$), while investors who received higher advice tended to suspect to be lied to by overstating ($L_{guess} > 0$).

Finding 2.6. The higher the received advice, the less the investors suspected to be lied to by understating and the more by overstating.

7.2.3. Summary

Table 9 summarizes the testing of my hypotheses that concern both players' sensitivity to in-game variables. Here, my predictions were fulfilled.

Table 9. Summary of sensitivity to in-game variables hypotheses testing.

Hypothesis	Concerned variables	Prediction	ρ-value (Spearman's rank correlation)	p-value (Spearman's rank correlation)	Result
H2.1	i^* and L	Negatively correlated	-0.985	<0.001	Supported
H2.2	a and L_{guess}	Positively correlated	0.773	<0.001	Supported
H2.3	a and $\bar{T}$ (with inverted sign)	Positively correlated	0.682	<0.001	Supported
H2.4	a and $\bar{T}_{guess}$ (with inverted sign)	Positively correlated	0.728	<0.001	Supported

On the one hand, the *advisors* reacted to different values of the optimal investment by adjusting their lying behavior. More precisely, the lower the given optimal investment, the more willing they were to overstate its value and, therefore, the higher was the extent to which they lied by overstating (see Finding 2.1; Hypothesis H2.1). In addition, the advisors had different expectations of their investors' mistrust depending on the given advice. The higher the advice, the less risk-seeking and the more risk-reducing mistrust they expected from their investors (see Finding 2.2; Hypothesis H2.4). In line with this, the advisors expected their investors to mistrust particularly high advice in such a way that giving higher advice above a certain amount would not lead to higher investments (see Finding 2.3).

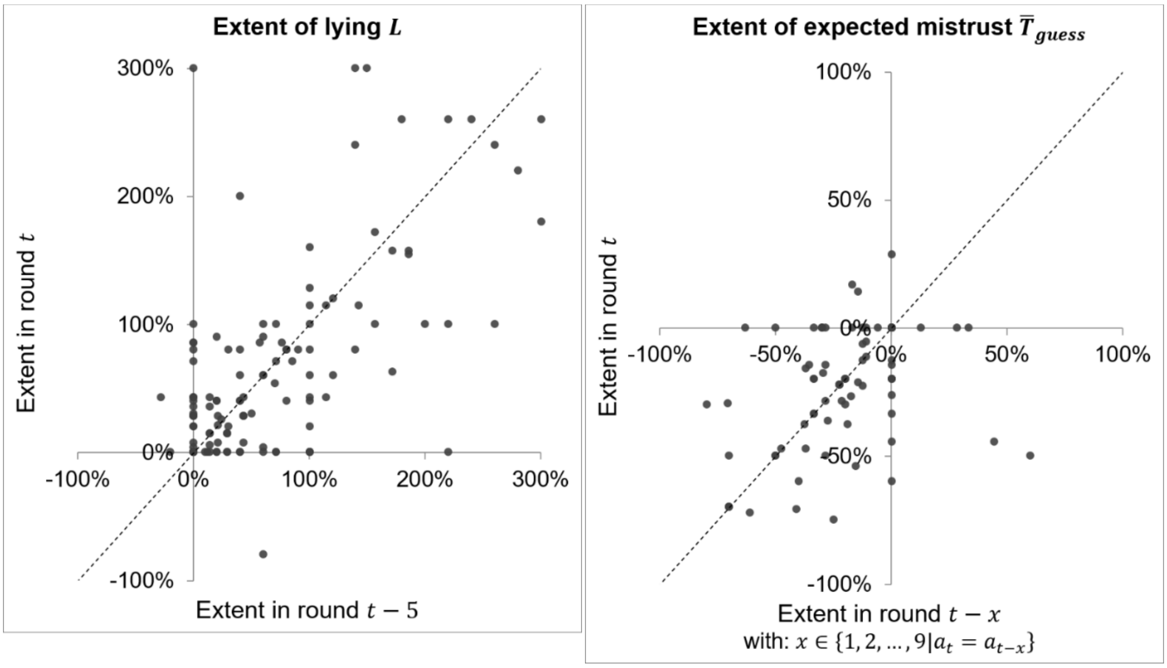
On the other hand, the *investors* adjusted both their behavior and first-order beliefs to their received advice. In particular, the higher their received advice, the less the investors engaged in risk-seeking and the more in risk-reducing mistrust (see Finding 2.4; Hypothesis H2.3). A similar pattern was observed regarding their beliefs about their advisors' lying behavior: The higher their received advice, the less they suspected their advisors to have lied by understating and the more by overstating (see Finding 2.6; Hypothesis H2.2). As expected by the advisors, the investors mistrusted particularly high advice by not following it above a certain level (see Finding 2.5).

7.3. Temporal consistency of honesty, trust, and beliefs

In this subsection I aim to test the consistency of both players' behavior and first-order beliefs about each other over time. As in the previous subsection, I begin with the advisors (*Lying and expected mistrust*) and then continue with the investors (*Mistrust and suspected lying*).

7.3.1. Lying and expected mistrust (advisors)

Figure 13 visualizes lag plots for the percentage extent of lying (L) on the left (13a) and the percentage extent of expected mistrust ($\overline{T}_{guess}$) on the right (13b).[68] Please note that, as shown before, the percentage extent of lying depends 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 13a) are lagged by five rounds. Turning to the plot on the right (Figure 13b), it should be reminded that the percentage extent of expected mistrust depends on the given advice. Therefore, to examine the temporal consistency of the advisors' beliefs about their investors' mistrust, it is reasonable to compare only rounds with identical advice. For that reason, the plot on the right (Figure 13b) 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.[69]



(**a**) (**b**)

Figure 13. 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 $\overline{T}_{guess}$.

The lag plot for the percentage extent of lying (L) in Figure 13a shows that the *advisors' lying behavior* was at least partially consistent over time, since the percentage extent of lying correlates significantly positively with its time-lagged values (Spearman's rank correlation with outlier-cleaned values: $\rho = 0.546$ with $p < 0.001$). On this basis, I find strong support for *hypothesis H3.1*. 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

[68] For the purpose of illustration, only the most relevant section of the plot in Figure 13a is displayed.

[69] Note that 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 $p = 0.250$).

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.[70] Moreover, advisors who lied ($L \neq 0$) the first time the optimal investment had the same value maintained the orientation of their behavior disproportionately more often over time than the other advisors (in 80.83% vs. 34.29% of cases; Fisher exact test: $p < 0.001$). This indicates that lying behavior was more consistent over time than honest behavior.

Finding 3.1. The advisors' lying behavior (L) was partially consistent over time. In particular, lying behavior was more consistent than honest behavior.

Turning to the lag plot for the percentage extent of expected mistrust ($\bar{T}_{guess}$) in Figure 13b, it can be seen that the *advisors' first-order beliefs* about their investors' mistrust were partially consistent over time, too. 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 $p < 0.001$), which is consistent with *hypothesis H3.3*. This is due to the fact that 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 if 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 about their investors' mistrust over time.[71] In particular, advisors who expected risk-reducing mistrust ($\bar{T}_{guess} < 0$) the first time they gave advice with the same value maintained the orientation of their beliefs disproportionately more often over time than the other advisors (in 77.59% vs. 53.13% of cases; Fisher exact test: $p = 0.031$). This indicates that expectations of risk-reducing mistrust were more consistent over time than different expectations.

Finding 3.2. The advisors' first-order beliefs about their investors' mistrust ($\bar{T}_{guess}$) were partially consistent over time. In particular, expectations of risk-reducing mistrust were more consistent over time than others.

7.3.2. Mistrust and suspected lying (investors)

Figure 14 displays lag plots for the percentage extent of mistrust ($\bar{T}$) on the left (14a) and the percentage extent of suspected lying (L_{guess}) on the right (14b).[72] Note that it has been shown before that both the percentage extent of mistrust and the percentage extent of suspected lying depend on the value of the received advice. To analyze the temporal consistency of the investors' behavior and first-order beliefs, it therefore is reasonable to compare only rounds with identical advice. For that reason, both plots in Figure 14 consider only investors who received the same advice at least twice. Thus, 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.[73]

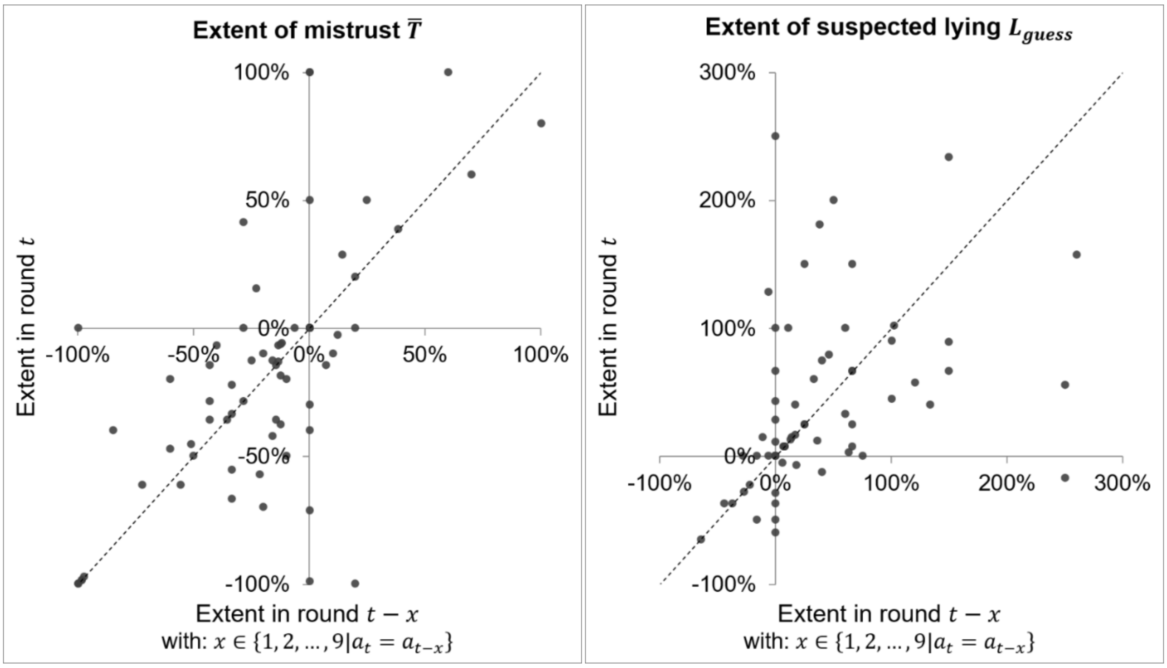
---

[70] This refers to whether they gave honest advice, lied by understating, or lied by overstating.

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

[72] For the purpose of illustration, only the most relevant section of the plot in Figure 14b is displayed.

[73] 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 $p = 0.358$).



(**a**) (**b**)

Figure 14. 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 14a reveals that the *investors' mistrusting behavior* was at least partially 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 $p < 0.001$). This offers strong support for *hypothesis H3.2*. Notably, 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, 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 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.[74] Moreover, investors who engaged in risk-reducing mistrust ($\bar{T} < 0$) the first time they received advice with the same value maintained the orientation of their behavior disproportionately more often over time than the other investors (in 88.37% vs. 63.64% of cases; Fisher exact test: $p = 0.013$). This indicates that strategies that involve risk-reducing mistrust were more consistent over time than strategies that do not.

Finding 3.3. The investors' mistrust ($\bar{T}$) was partially consistent over time. In particular, risk-reducing mistrust was more consistent over time than other behaviors.

It can be read from the lag plot of the percentage extent of suspected lying (L_{guess}) in Figure 14b that the *investors' first-order beliefs* about their advisors' lying behavior were partially consistent over time, too. This is because the percentage extent of suspected lying correlates significantly positively with its time-lagged values (Spearman's rank correlation with outlier-cleaned values: $\rho = 0.597$ with $p < 0.001$). Thus, I find strong support for *hypothesis H3.4*. 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

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

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 about their advisors' lying behavior over time.[75] In particular, investors who suspected a piece of advice to be a lie ($L_{guess} \neq 0$) the first time they received it maintained the orientation of their beliefs disproportionately more often over time than the other investors (in 81.82% vs. 52.38% of cases; Fisher exact test: p = 0.018). This suggests that expectations of being lied to were more consistent over time than expectations of being told the truth.

Finding 3.4. The investors' first-order beliefs about their advisors' lying behavior (L_{guess}) were partially consistent over time. In particular, expectations of being lied to were more consistent over time than expectations of being told the truth.

7.3.3. Summary

Table 10 provides an overview of the testing of my hypotheses about the temporal consistency of both players' behavior and first-order beliefs.

The *advisors*' lying behavior (see Finding 3.1; Hypothesis H3.1) and their expectations of being mistrusted by their investors (see Finding 3.2; Hypothesis H3.3) were both partially consistent over time. This was because most advisors neither changed the orientation of their behavior nor of their beliefs about the investors over time. In particular, advisors who lied maintained the orientation of their behavior disproportionately more often than advisors who told the truth (see Finding 3.1). Moreover, advisors who expected risk-reducing mistrust changed the orientation of their first-order beliefs disproportionately less often over time than other advisors (see Finding 3.2). It can be concluded that strategies that involve lying or expectations of risk-reducing mistrust were more consistent over time than other strategies.

Table 10. Summary of temporal consistency hypotheses testing.

Hypothesis	Concerned variables	Prediction	ρ-value (Spearman's rank correlation)	p-value (Spearman's rank correlation)	Result
H3.1	L and time-lagged L	Positively correlated	0.546	<0.001	Supported
H3.2	$\bar{T}$ and time-lagged $\bar{T}$	Positively correlated	0.627	<0.001	Supported
H3.3	$\bar{T}_{guess}$ and time-lagged $\bar{T}_{guess}$	Positively correlated	0.482	<0.001	Supported
H3.4	L_{guess} and time-lagged L_{guess}	Positively correlated	0.597	<0.001	Supported

The *investors*' mistrusting behavior (see Finding 3.3; Hypothesis H3.2) and their expectations of being lied to by their advisors (see Finding 3.4; Hypothesis H3.4) were both partially consistent over time. That can be explained by the fact that most investors neither changed the orientation of their behavior nor of their beliefs about their advisors over time. In particular, investors who engaged in risk-reducing mistrust maintained the orientation of their behavior disproportionately more often than other investors (see Finding 3.3). Moreover, investors who expected a piece of advice to be a lie the first time they received it changed these beliefs disproportionately less often over time than investors who

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

expected it to be truthful (see Finding 3.4). In short, strategies that involve risk-reducing mistrust or expectations of being lied to were more consistent over time than other strategies.

7.4. Strategy analysis: the relationship between honesty, trust, and beliefs

In this subsection I examine which strategies both players pursued in the Continuous Deception Game. Firstly, I will analyze the short-term relationship between players' behavior and first-order beliefs about each other. Secondly, I will check for tendencies in their strategic orientation. I will begin with the advisors (*Lying and expected mistrust*) and then turn to the investors (*Mistrust and suspected lying*).

7.4.1. Lying and expected mistrust (advisors)

Figure 15 visualizes the relationship between the advisors' lying behavior and their first-order beliefs about their investors.[76] 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 between L and $-\bar{T}_{guess}$ with outlier-cleaned values: $\rho = 0.354$ with $p < 0.001$). This relationship provides support for *hypothesis H4.1*. 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 different 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.

Finding 4.1. The advisors' lying behavior (L) is closely related to their expectations of being mistrusted ($\bar{T}_{guess}$).

[76] For the purpose of illustration, only the most relevant area of the plot in Figure 15 is displayed. Please also note that the dotted line in Figure 15 marks the hypothetical line on which the advisors expected the investments to be optimal (which would imply: $F_{A;guess} = 0$). Hence, points below this line represent *expectations of underinvestments* ($F_{A;guess} < 0$), whereas points above it represent *expectations of overinvestments* ($F_{A;guess} > 0$).

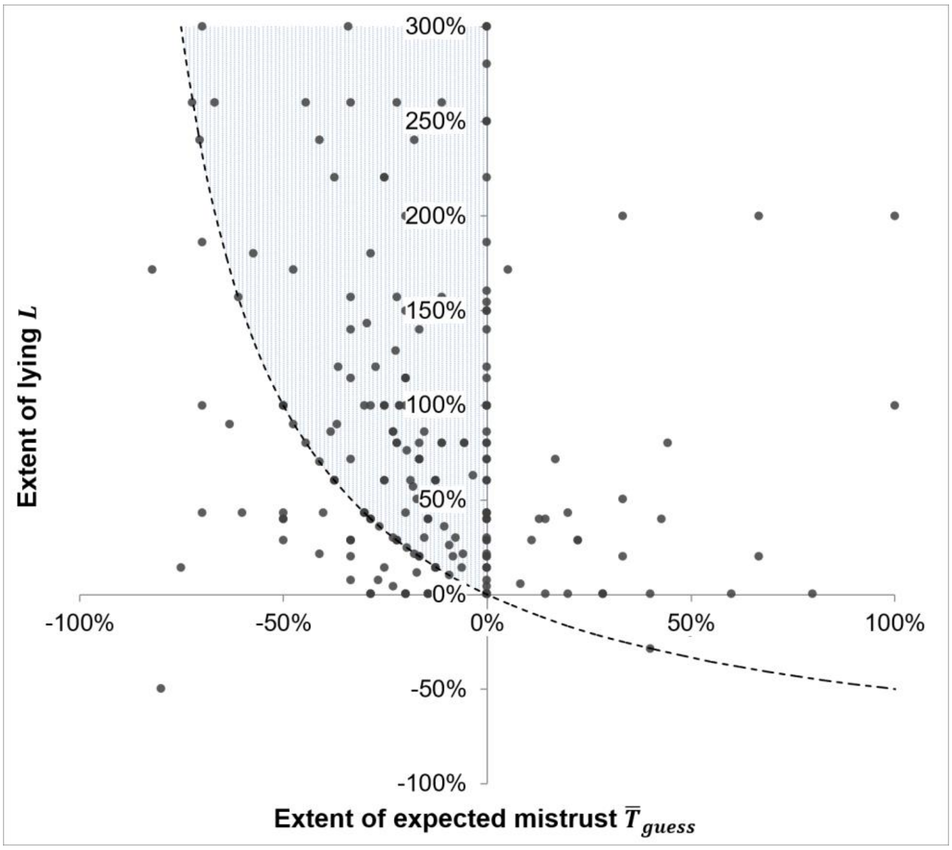


Figure 15. Relationship between the extent of lying L and the extent of expected mistrust $\bar{T}_{guess}$.

To analyze the *effective direction* between these two variables, I will examine how they affected each other over time. For this purpose, I will check to what extent the advisors' lying behavior predicted their expectations about their investors' mistrust over time (Direction 1: Lying by overstating → Expected risk-reducing mistrust) – and the other way around (Direction 2: Expected risk-reducing mistrust → Lying by overstating). Based on this, I will estimate the effective direction of the round-lagged short-term[77] relationship between lying by overstating and expectations of risk-reducing mistrust (L and $-\bar{T}_{guess}$). Therefore, I will compare the corresponding round-lagged correlations of both directions to one another. The time lag that I will use here consists of five rounds, since this will ensure that only rounds with identical optimal investment will be matched.

Starting with direction 1 (Lying by overstating → Expected risk-reducing mistrust), the correlation between the advisors' percentage extent of lying by overstating (L) before the time lag and their expectations of risk-reducing mistrust ($\bar{T}_{guess}$ with inverted sign) after the time lag is weak and not significant (Spearman's rank correlation between L and $-\bar{T}_{guess}$ with outlier-cleaned values: $\rho = 0.058$ with $p = 0.519$). It can be assumed that the advisors did not adjust their first-order beliefs about their investors' mistrust to their lying behavior over the time lag. However, the picture changes for direction 2 (Expected risk-reducing mistrust → Lying by overstating). Here, the advisors' percentage extent of expected risk-reducing mistrust ($\bar{T}_{guess}$ with inverted sign) before the time lag is significantly positively correlated with their lying behavior (L) after the time lag (Spearman's rank correlation between L and $-\bar{T}_{guess}$ with outlier-cleaned values: $\rho = 0.325$ with $p < 0.001$). This suggests that the advisors' beliefs about their investors' mistrust influenced their lying behavior over time. Comparing both directions shows that the round-lagged correlation for direction 2 ($-\bar{T}_{guess} \rightarrow L$) is significantly stronger than the correlation for direction 1 ($L \rightarrow -\bar{T}_{guess}$) (two-sided Fisher's Z test: $p = 0.028$). Hence, I find support for *hypothesis H4.2*. It can be concluded that lying by overstating was preceded by expectations of risk-

[77] Note that each round took less than 90 seconds to complete.

reducing mistrust and not the other way around. This indicates that the advisors adapted their behavior to their first-order beliefs about the investors' behavior over time, which suggests a tendency for strategic decision making.

Finding 4.2. The advisors' expectations of being mistrusted by their investors ($\bar{T}_{guess}$) predicted their lying behavior (L) over time – rather than the other way around.

To check for tendencies in the strategic orientation of the advisors, Table 11 summarizes the proportions of *classes of observed advisor strategies* (which are also displayed in a non-aggregated form in Figure 15). This summary is based on the aforementioned taxonomy of lies and truth-telling.

Table 11. Proportions of advisor strategy classes.

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

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*, i.e., game theoretical equilibrium strategies, in 77.74% of all cases.[78] In Figure 15 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

[78] This means that in each of these cases exists at least one function for the advisor's moral costs of lying $C_L(a)$ that could rationally explain their behavior from a game theoretical perspective. Hence, with this function their pursued strategy would be an equilibrium strategy.

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.[79] It comes as no surprise that the advisors pursued these potential equilibrium strategies more often than other strategies in every round (two-sided Binomial tests: $p < 0.001$ for each round[80]). This is consistent with *hypothesis H4.3* and therefore supports the idea of rational decision making based on individual lie aversion and rational beliefs.

Finding 4.3. The advisors pursued rational strategies that correspond to potential equilibrium strategies disproportionately more often (in 77.74% of cases) than other strategies. In most of all cases (57.74%) the advisors were selfish liars.

7.4.2. Mistrust and suspected lying (investors)

The scatter plot in Figure 16 visualizes the relationship between the investors' mistrusting behavior and their first-order beliefs about their advisors.[81] 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} with outlier-cleaned values: $\rho = 0.738$ with $p < 0.001$). Hence, I find strong support for *hypothesis H4.4*. 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 different 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.

Finding 4.4. The investors' mistrusting behavior ($\bar{T}$) is closely related to their expectations of being lied to by their advisors (L_{guess}).

[79] Note that "only" 89.39% of all selfish liars pursued a potential equilibrium strategy, since the rest of them had beliefs about their investors' behavior that are not rational from a game theoretical point of view.

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

[81] For the purpose of illustration, only the most relevant area of the plot in Figure 16 is displayed. Please also note that the dotted line in Figure 16 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* ($F_{I;guess} > 0$).

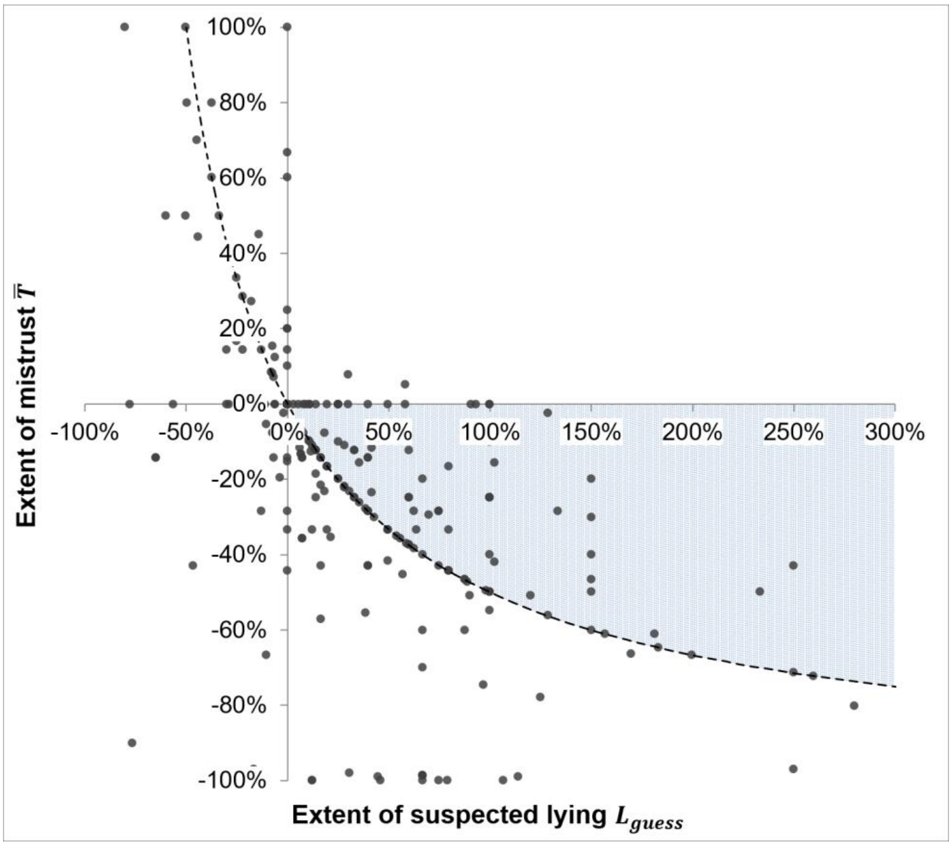


Figure 16. Relationship between the extent of mistrust $\bar{T}$ and the extent of suspected lying L_{guess}.

In order to draw conclusions about the *effective direction* between these two variables, I will analyze how they affected each other over time. Therefore, I will examine to what extent the investors' mistrusting behavior predicted their expectations about their advisors' lying behavior over time (Direction 1: Risk-reducing mistrust → Suspected lying by overstating) – and the other way around (Direction 2: Suspected lying by overstating → Risk-reducing mistrust). On this basis, I will estimate the effective direction of the round-lagged short-term[82] relationship between risk-reducing mistrust and expectations of lying by overstating ($-\bar{T}$ and L_{guess}). For this purpose, I will compare the corresponding round-lagged correlations of both directions to one another. Since it has been shown before that both variables depend on the given advice, I will only consider investors who received the same advice at least twice. Thus, the time lag 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.[83]

On the one hand, for direction 1 (Risk-reducing mistrust → Suspected lying by overstating), the investors' percentage extent of risk-reducing mistrust ($\bar{T}$ with inverted sign) before the time lag is significantly positively correlated with their expectations of overstating lies (L_{guess}) after the time lag (Spearman's rank correlation between $-\bar{T}$ and L_{guess} with outlier-cleaned values: $\rho = 0.434$ with $p < 0.001$). This could suggest that the investors adjusted their beliefs about their advisors to their own prior mistrusting behavior. On the other hand, for direction 2 (Suspected lying by overstating → Risk-reducing mistrust), the correlation between the investors' percentage extent of suspected lying (L_{guess}) before the time lag and their percentage extent of risk-reducing mistrust ($\bar{T}$ with inverted sign) after the time lag is strong and significant, too (Spearman's rank correlation between $-\bar{T}$ and L_{guess} with outlier-cleaned values: $\rho = 0.607$ with $p < 0.001$). This indicates that risk-reducing mistrust was preceded by

[82] Note that each round took less than 90 seconds to complete.

[83] 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 $p = 0.358$).

expectations of overstating lies. The comparison between both directions reveals that the round-lagged correlation is moderately but barely non-significantly stronger for direction 2 ($L_{guess} \rightarrow -\bar{T}$) than for direction 1 ($-\bar{T} \rightarrow L_{guess}$) (two-sided Fisher's Z test: p = 0.157). On this basis, I find only weak support for *hypothesis H4.5*. Even though this does not clearly speak for or against one effective direction, the observed difference between both correlations suggests that risk-reducing mistrust can be predicted by expectations of overstating lies better than the other way around. This in turn would indicate that the investors adapted their behavior to their first-order beliefs about their advisors' behavior over time, which would suggest a slight tendency for more strategic rather than intuitive decision making.

Finding 4.5. The investors' expectations of being lied to by their advisors (L_{guess}) predicted their mistrusting behavior ($\bar{T}$) over time moderately, even though barely non-significantly, better than the other way around.

To illustrate the tendencies in the strategic orientation of the investors, Table 12 provides an overview of the proportions of *classes of observed investor strategies* (which are also displayed in a non-aggregated form in Figure 16). This summary is based on the earlier introduced taxonomy of mistrust and trust.

Table 12. Proportions of investor strategy classes.

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

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.[84] 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

[84] 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 16, this refers to all strategy points that are located on the dotted line except for the ones at the point of origin.

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.

Assuming that the investors' behavior was consistent with their beliefs and their individual preferences for trust, they pursued *rational strategies*, i.e., game theoretical equilibrium strategies, in 60.00% of all cases.[85] In Figure 16 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 potential equilibrium strategies more often than other strategies (two-sided Binomial test[86]: $p < 0.001$). This finding is consistent with *hypothesis H4.6*, which supports the idea of rational decision making based on individual trust preferences and rational beliefs.

Finding 4.6. The investors pursued rational strategies that correspond to potential equilibrium strategies disproportionately more often (in 60.00% of cases) than other strategies. Most of them engaged in profitable mistrust.

7.4.3. Summary

Tables 13a, b, and c summarize the testing of my hypotheses regarding both players' strategic orientation based on the relationship between their behavior and first-order beliefs.

Table 13a. Summary of strategy hypotheses testing.

Hypothesis	Concerned variables	Prediction	ρ-value (Spearman's rank correlation)	p-value (Spearman's rank correlation)	Result
H4.1	$\bar{T}_{guess}$ (with inverted sign) and L	Positively correlated	0.354	<0.001	Supported
H4.4	L_{guess} and $\bar{T}$ (with inverted sign)	Positively correlated	0.738	<0.001	Supported

The *advisors* pursued mostly rational, i.e., potential equilibrium, strategies (see Finding 4.3; Hypothesis H4.3). In particular, most of them engaged in selfish lying. Overall, their lying behavior was closely related to their expectations of being mistrusted (see Finding 4.1; Hypothesis H4.1). In fact, lying by overstating the value of the optimal investment was preceded by expectations of risk-reducing

[85] This means that in each of these cases exists at least one function for the investor's costs of mistrust $C_{\bar{T}}(a, i)$ that could rationally explain their behavior from a game theoretical perspective. Hence, with this function their pursued strategy would be an equilibrium strategy.

[86] The probability $p_{I;rat}$ that an investor would engage in 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: $p_{I;rat}(a) = 0.5 * \left(\frac{a}{i_{max}}\right)^2$.

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, this upper limit of $p_{I;rat}$ was used to perform the most conservative two-sided Binomial test possible that compares the proportion of potential equilibrium strategies to its expected value based on random choices.

mistrust – but not the other way around (see Finding 4.2; Hypothesis H4.2). These findings speak in favor of rational and strategic decision making.

Table 13b. Summary of strategy hypotheses testing.

Hypothesis	Concerned variables	Prediction	p-value (two-sided Fisher's Z test[1])	Result
H4.2	Time-lagged correlations between $\bar{T}_{guess}$ and L	Direction ($\bar{T}_{guess} \rightarrow L$) > Direction ($L \rightarrow \bar{T}_{guess}$)	0.028	Supported
H4.5	Time-lagged correlations between L_{guess} and $\bar{T}$	Direction ($L_{guess} \rightarrow \bar{T}$) > Direction ($\bar{T} \rightarrow L_{guess}$)	0.157	Weakly supported

[1] Fisher's Z test to compare both time-lagged correlations. For more details, see Fisher [56].

The *investors*' mistrusting behavior was closely related to their expectations of being lied to by their advisors (see Finding 4.4; Hypothesis H4.4). In addition, their expectations of being lied to predicted their mistrusting behavior over time moderately and just barely non-significantly better than the other way around (see Finding 4.5; Hypothesis H4.5). Moreover, the majority of investors pursued rational, i.e., potential equilibrium, strategies, most of them engaging in profitable mistrust (see Finding 4.6; Hypothesis H4.6). Even though intuitive decision making cannot be completely ruled out by these findings, all things considered, this speaks in favor of rational and strategic decision making.

Table 13c. Summary of strategy hypotheses testing.

Hypothesis	Concerned variables	Prediction	p-value (two-sided Binomial test[1])	Result
H4.3	#rational strategies and #non-rational strategies (*pursued by advisors*)	#rational strategies > #non-rational strategies	<0.001[2]	Supported
H4.6	#rational strategies and #non-rational strategies (*pursued by investors*)	#rational strategies > #non-rational strategies	<0.001[3]	Supported

[1] Binomial test to compare the proportion of potential equilibrium strategies to its expected value based on random choices; [2] Since the optimal investment varied between rounds, a separate binomial test was performed for each round; [3] Since not all investors received the same piece of advice, the binomial test was based on the maximum probability that an investor would engage in a potential equilibrium strategy by making random choices.

7.5. Consistency between players' behavior and their self-assessment

In this subsection I will test how consistent my interpretation of both players' strategies in the Continuous Deception Game 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 (*Lying behavior*) with their self-assessed preference for risk and honesty. Then, I will examine how consistent the observed behavior of the investors (*Mistrust*) is with their self-assessed preference for risk and trust.

To begin with, 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.

In addition, 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.

Moreover, 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.

Please note that in the following evaluation of the questionnaire data all 7-point-scales are coded from 0 (lowest) to 6 (highest).

7.5.1. 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 5.1. The higher the advisors' percentage extent of lying (L), the more risk-seeking they evaluated their own behavior.

In the second place, the advisors' self-assessment of their *honesty* in the game did not differ significantly from their assessment of the honesty of the other advisors (ratings: 1.83 vs. 1.57; two-sided Wilcoxon signed-rank test: $p = 0.453$). This indicates that their self-assessed honesty was consistent in relation to the group. It comes as no surprise that the advisors' self-assessed honesty is significantly negatively correlated with the absolute value of the percentage extent to which they lied ($|L|$)[87] 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.

[87] Here, I use the absolute value of the percentage extent of lying ($|L|$), since the advisors' self-assessment of their 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.

Finding 5.2. The advisors' lying behavior (L) and their pursued strategies based on the taxonomy of lies and truth-telling are largely consistent with the advisors' self-assessment of their honesty.

7.5.2. 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 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 ex ante 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 5.3. The inherent risk of the investors' mistrusting behavior ($\bar{T}$) is consistent with their self-assessment of their preference for risk.

Turning to the investors' self-assessment of their *trust* in the game reveals that their evaluation of their own trust barely differed from their assessment of the trust of the other investors (ratings: 2.45 vs. 2.58; two-sided Wilcoxon signed-rank test: $p = 0.430$). Thus, their self-assessed trust was consistent in relation to the group. In addition, the investors' self-assessed trust correlates significantly negatively with the absolute value of the percentage extent to which they engaged in mistrusting behavior ($|\bar{T}|$)[88] 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 5.4. The investors' mistrust ($\bar{T}$) and their pursued strategies based on the taxonomy of mistrust and trust are highly consistent with the investors' self-assessment of their trust.

7.5.3. Summary

My ex ante classification of both players' strategies is largely consistent with their ex post self-assessment of their own behavior.

On the one hand, the *advisors*' lying behavior is strongly consistent with their ex post assessment of their own honesty (see Finding 5.2). Moreover, they considered truth-telling and cooperative behavior as honest, while considering selfish lying as dishonest. This is also consistent with my taxonomy of lies and truth-telling. Interestingly, a preference for risk was associated with dishonest behavior, too (see Finding 5.1).

On the other hand, the *investors*' mistrust is largely consistent with their ex post evaluation of their own trust (see Finding 5.4). In particular, they considered following their received advice in fact as trusting, which is consistent with my taxonomy of mistrust and trust. Moreover, my ex ante 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 (see Finding 5.3).

In the next section, I will discuss my findings against the backdrop of the existing literature and draw some conclusions about application possibilities of this new experimental design.

---

[88] Here, I use the absolute value of the percentage extent of mistrust ($|\bar{T}|$), since the investors' self-assessment of their trust did not differentiate between risk-reducing and risk-seeking mistrust. However, both of these types of mistrust can be considered as mistrusting behavior.

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.*[89] This experiment is framed as an investment game and features two players: an advisor (i.e., the sender) with complete and an investor (i.e., the receiver) with incomplete information. On the one hand, the advisor is incentivized to lie by overstating the true value of an optimal investment, which he or she is instructed to report to the investor. On the other hand, the investor is incentivized to invest optimally and, therefore, must rely on the alleged optimum reported by the advisor. In addition, both players' first-order beliefs are elicited. As a result, lying, mistrust and respective first-order beliefs are observed on continuous scales. This enables testing of more sophisticated theoretical predictions.

I find that both players tended to make strategic and belief-based decisions. In particular, they adjusted their behavior to their beliefs about the other player over time and pursued rational (i.e., potential equilibrium) strategies disproportionally more often than other ones (*section 7.4.*). As a consequence, both players' financial success in the game was largely determined by the accuracy of their first-order beliefs (*section 7.1.*). Here, the advisors took advantage of the fact that their first-order beliefs were much more accurate than those of the investors. On this basis, most of the advisors successfully manipulated their investors into overinvesting. This result can be explained by the fact that the investors based their investments too much on their received advice. In fact, both players' strategies were strongly based on their individual information about the value of the optimal investment (*section 7.2.*). Therefore, with identical information input, their behavior and expectations of each other were relatively consistent over time (*section 7.3.*). Finally, the analysis of my post-experimental questionnaire shows that my interpretation of both players' strategies is largely consistent with their ex post self-assessment of their own behavior within the experiment (*section 7.5.*).

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 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.[90] 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%. As mentioned earlier, 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 extents of lying and mistrust into the picture when one pursues to understand how dishonest or mistrusting the players actually behaved.

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, respectively, 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 (*Lying and expected mistrust*) and, secondly, for the investors (*Mistrust and suspected lying*).

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 [6] 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 [6] varies possible gains and losses from lying between treatments, in my experiment the senders (i.e.,

[89] The experimental design is largely inspired by Erat and Gneezy's [7] *Deception Game*, which originated from Gneezy [6].

[90] Both proportions were approximately 78%.

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.[91] Transferring Gneezy's [6] 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 [9] and Gneezy et al. [10] 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 [10,13]. They are also in line with the aim to maintain a favorable self-concept [14]. 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 [55].

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 mistrust particularly high advice. As my results generally speak in favor of highly strategic and belief-based decision making, this indicates that some advisors simply thought that lying to a higher extent than they already did would not deceive their investors any further and, thus, not yield higher payoffs. In other words, my findings suggest that people who refrain from lying to the fullest possible extent might still lie to the highest extent they expect to be convincing. Such considerations about the own *potential to manipulate* in strategic deception cannot be addressed in most other economic experiments, since lies in other experiments often do not have to be convincing in order to yield favorable results for the liar (e.g., [9,14,55]).[92]

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 [13] 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 size of the senders' lies did not matter beyond the fact if 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. [13] 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

[91] Please note that Gneezy [6] and I use the same ratio between the profits from lying to the senders and the associated costs to the receivers. When Gneezy [6] 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 [6].

[92] I argue that this also applies to all sender-receiver games that feature binary-like choices (e.g., [6,7,12,37-45]). While such experiments may allow for sophisticated truth-telling (as shown by Sutter [39]), considerations about the own potential to manipulate the other player are still strictly limited by binary-like strategy sets.

person provide a reasonable explanation for their results – namely that the senders in Lundquist et al.'s [13] 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. Therefore, it seems highly important to consider this aspect when comparing different-sized lies in particular and when analyzing the intent behind lying or any other deceptive behavior in general.[93]

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 [7] 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. In addition, 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 and Kartik [11]). This indicates that the respective players were dealing with moral costs of lying that increase with the size of the lie (as suggested by Lundquist et al. [13] and Gneezy et al. [10]).

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 and Gneezy [7]), which is in contrast to Vanberg [57] who finds no evidence for the existence of such a motivation in his experiment. On the other hand, the low 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 [57], as this means that intrinsic lie aversion alone seems to not have been a strong driver for absolute truth-telling in my experiment.

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

Truth-telling is generally seen in a more positive light than telling a lie. While this assessment is certainly true in most cases, not all lies are of bad intentions. For instance, in Erat and Gneezy's [7] sender-receiver game a significant fraction of senders chose to engage in altruistic white lies (i.e., lies

[93] Gneezy et al. [10] 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, they 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 solely asked players 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, while second-order beliefs do not provide much more insight [12]. While this allows keeping my experiment simpler, it makes it hard to ex post distinguish social identity concerns [10] from strategic considerations about one's potential to manipulate others. To make this distinction, one could conduct a version of my Continuous Deception Game 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 Gneezy [6], Mazar et al. [14], Sutter [39], Erat and Gneezy [7], Fischbacher and Föllmi-Heusi [9], and Pruckner and Sausgruber [16]), 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.

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 their extent of lying 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 [7], 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 [7] 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 [7] 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 [7] 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 extent of lying. 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., [9,14,16]) or at least focus more on the former than on the latter (e.g., [6,7,12,13,44]). While this focus on lying behavior serves the purpose of these studies in a good way, it also ignores the importance of trust for economic decision making. To address this issue, my experiments allows drawing conclusions on why people trust or mistrust potential liars.[94]

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 [41] and the Trust Game [29] 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

[94] Please note that, in my experiment, *trust* refers to honesty-related trusting behavior (i.e., one's reliance on another person's honesty), while *trusting beliefs* correspond to expectations about the honesty of another person. In this respect, my understanding of trust differs from that in the original version of the Trust Game in which trust refers to expectations of social cooperation.

way, they raised their expected payoffs. For this reason, I refer to such behavior as profitable mistrust, or rather profitable white mistrust if it increased the advisors' payoffs, too. Similar to cooperative trust, profitable (white) mistrust could be motivated by *monetary gain*.

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 by design in my experiment.[95] 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. [29] 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 my experiment).

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. [29] who observe a similar pattern for trusting behavior in the Trust Game. In addition, they are consistent with Cox [58] and Innocenti and Pazienza [59] 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 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 [25,27] and mistrust to negative sensations [28].

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

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

9. Conclusions

By introducing continuous variables to the sender-receiver game, the *Continuous Deception Game* 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 [39]. 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, this experiment 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 [7] 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 when lying. 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 Continuous Deception Game. This demonstrates the variety of application possibilities for this experiment and shows its potential as a straightforward method to analyze the relationship between honesty, trust, and respective beliefs in strategic deception.

Appendix A. Derivation of the investor's expected payoff and best response (with only monetary motivation)

A.1. Derivation of the investor's expected payoff

In this appendix, I derive the investor's expected payoff $\pi^e_{I;hoec}(i)$ with players who are purely interested in their monetary payoff and, thus, have no preference for honesty or trust. I start by analyzing the investor's payoff structure: Given a predefined maximal investment $i_{max} > 0$ and an optimal investment $i^* \in [0, i_{max}]$, which is randomly determined by a uniform distribution ($P([0, i^*]) = \frac{i^*}{i_{max}}$ with $i^* \in [0, i_{max}]$), the investor's payoff $\pi_I(i)$ is defined as the following function of the investment $i \in [0, i_{max}]$:

$$\pi_I(i) = \begin{cases} i \leq i^*: & i_{max} + m_I * i \\ i > i^*: & i_{max} + m_I * (2 * i^* - i) \end{cases} \tag{25}$$

with $m_I > 0$.

This function is designed in such a way that it is maximized by the optimal investment i^*, which is not known to the investor. Furthermore, this type of investor has no reason to trust the advice a and, therefore, disregards this information. Hence, he or she assumes that the true optimal investment i^* is located somewhere between 0 and the maximal investment i_{max} with equal probability.

Hence, the investor's expected payoff $\pi^e_{I;hoec}(i)$ can be defined as the following function of the investment i:

$$\pi^e_{I;hoec}(i) = \frac{1}{i_{max}} \int_0^{i_{max}} \pi_I(i)\, di^*. \tag{26}$$

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:

$$\begin{aligned}
\pi^e_{I;hoec}(i) &= \frac{1}{i_{max}} \int_0^{i_{max}} \pi_I(i)\, di^* \\
&= \frac{1}{i_{max}} * \lim_{b \nearrow i} \left(\int_0^b \underbrace{\pi_I(b)}_{i^*<i} di^* \right) + \frac{1}{i_{max}} \int_i^{i_{max}} \underbrace{\pi_I(i)}_{i^* \geq i} di^* \\
&= \frac{1}{i_{max}} * \lim_{b \nearrow i} \left(\int_0^b \big(i_{max} + m_I * (2 * i^* - b)\big)\, di^* \right) + \frac{1}{i_{max}} \int_i^{i_{max}} (i_{max} + m_I * i)\, di^* \\
&= \frac{1}{i_{max}} \int_0^i \big(i_{max} + m_I * (2 * i^* - i)\big)\, di^* + \frac{1}{i_{max}} \int_i^{i_{max}} (i_{max} + m_I * i)\, di^* \\
&= \frac{1}{i_{max}} * \left[i_{max} * i^* + m_I * \left(i^{*2} - i * i^*\right)\right]_0^i + \frac{1}{i_{max}} * [(i_{max} + m_I * i) * i^*]_i^{i_{max}} \\
&= \frac{1}{i_{max}} * [i_{max} * i + m_I * (i^2 - i^2)] + \frac{1}{i_{max}} * (i_{max} - i) * (i_{max} + m_I * i) \\
&= \frac{i}{i_{max}} * i_{max} + \left(1 - \frac{i}{i_{max}}\right) * (i_{max} + m_I * i) \\
&= i_{max} + m_I * \left(i - \frac{i^2}{i_{max}}\right)
\end{aligned} \tag{27}$$

with $m_I > 0$.

A.2. Derivation of the investor's best response

In this appendix, I determine the investor's best response $i_{BR;hoec}$ with players who are purely interested in their monetary payoff and, therefore, have no preference for honesty or trust. I start with the previously derived function for this type of investor's expected payoff $\pi^e_{I;hoec}(i)$. Given a maximal investment $i_{max} > 0$, it is defined by the following function of the investment $i \in [0, i_{max}]$:

$$\pi^e_{I;hoec}(i) = i_{max} + m_I * \left(i - \frac{i^2}{i_{max}}\right) \quad (28)$$

with $m_I > 0$.

Furthermore, this type of investor's best response $i_{BR;hoec}$ is defined as the investment i that maximizes their expected payoff $\pi^e_{I;hoec}(i)$. It follows:

$$\begin{aligned}
\max_i\left(\pi^e_{I;hoec}(i)\right) \quad &\rightarrow \quad \frac{\partial \pi^e_{I;hoec}(i)}{\partial i} \overset{!}{=} 0 \\
&\leftrightarrow \quad \frac{\partial\left(i_{max} + m_I * \left(i - \frac{i^2}{i_{max}}\right)\right)}{\partial i} = 0 \\
&\leftrightarrow \quad m_I - \frac{2 * m_I * i}{i_{max}} = 0 \\
&\leftrightarrow \quad \frac{2 * i}{i_{max}} = 1 \\
&\leftrightarrow \quad i_{BR;hoec} = \frac{i_{max}}{2}
\end{aligned} \quad (29)$$

with $m_I > 0$.

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

$$\begin{aligned}
\pi^e_{I;BR;hoec} &= \pi^e_{I;hoec}\left(i = i_{BR;hoec}\right) \\
&= i_{max} + m_I * \left(i_{BR;hoec} - \frac{{i_{BR;hoec}}^2}{i_{max}}\right) \\
&= i_{max} + m_I * \left(\frac{i_{max}}{2} - \left(\frac{i_{max}}{2}\right)^2 * \frac{1}{i_{max}}\right) \\
&= i_{max} * \left(1 + \frac{m_I}{4}\right)
\end{aligned} \quad (30)$$

with $m_I > 0$.

It should be noted that this type of investor's best response $i_{BR;hoec}$ is always half of the maximal investment i_{max} and, thus, does not depend on the received advice a ($\forall a$: $i = i_{BR;hoec} = \frac{i_{max}}{2}$).

Appendix B. Derivation of set of game theoretical equilibria (with monetary and non-monetary motivation)

In this appendix, I specify the set of game theoretical equilibria of the Continuous Deception Game with rational players who consider honesty and trust besides their monetary payoffs.

The following is given:

- i_{max}: maximal investment,
- i^*: optimal investment: This investment is randomly determined by a uniform distribution ($P([0, i^*]) = \frac{i^*}{i_{max}}$ with $i^* \in [0, i_{max}]$).
- a: advice with: $a \in [0, i_{max}]$,
- i: investment with: $i \in [0, i_{max}]$,
- i_{guess}: advisor's guess about the investment i with: $i_{guess} \in [0, i_{max}]$,
- i^*_{guess}: investor's guess about the optimal investment i^* with: $i^*_{guess} \in [0, i_{max}]$,
- $\pi_A(i)$: advisor's monetary payoff function with: $\pi_A(i) = m_A * i$ with $m_A > 0$,
- $\pi_I(i)$: investor's monetary payoff function with: $\pi_I(i) = \begin{cases} i \leq i^*: & i_{max} + m_I * i \\ i > i^*: & i_{max} + m_I * (2 * i^* - i) \end{cases}$ with $m_I > 0$,
- L: percentage extent of lying with: $L = \frac{a - i^*}{i^*}$ with $i^* > 0$,
- L_{guess}: percentage extent of suspected lying with: $L_{guess} = \frac{a - i^*_{guess}}{i^*_{guess}}$ with $i^*_{guess} > 0$,
- $\bar{T}$: percentage extent of mistrust with: $\bar{T} = \frac{i - a}{a}$ with $a > 0$,
- $\bar{T}_{guess}$: percentage extent of expected mistrust with: $\bar{T}_{guess} = \frac{i_{guess} - a}{a}$ with $a > 0$,
- $C_L(a)$: advisor's moral costs of lying: These costs are a function of the advice a in relation to the predefined optimal investment i^*. They represent the advisor's preference for honesty. The exact nature of this function depends on the advisor's type.[96]
- $C_{\bar{T}}(a, i)$: 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 depends on the investor's type.[97]
- $U_A(a, i)$: 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: $U_I(a, i) = \pi_I(i) - C_{\bar{T}}(a, i)$.

It is assumed that both players aim to maximize their utility ($U_A(a, i)$ or, respectively, $U_I(a, i)$). Based on this, I will first specify (*B.1.*) the advisor's set of best responses $S_{A;BR}$. Secondly, I will derive (*B.2.*) the investor's set of best responses $S_{I;BR}$. Thirdly, I will specify (*B.3.*) the set of game theoretical equilibria S_{equ}. Finally, I will outline (*B.4.*) some implications for the impact of both players' beliefs on their behavior in equilibrium.

---

[96] 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 $\left(|L| = \left|\frac{a - i^*}{i^*}\right|\right)$: $\frac{\partial(C_L(a))}{\partial a} = \begin{cases} a < i^*: & \leq 0 \\ a > i^*: & \geq 0 \end{cases}$.

[97] The investor's costs of mistrust $C_{\bar{T}}(a, i)$…

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

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

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

B.1. The advisor's set of best responses $S_{A;BR}$

To begin with, the advisor aims to maximize their utility

$$\begin{aligned} U_A(a,i) &= \pi_A(i) - C_L(a) \\ &= m_A * i - C_L(a) \end{aligned} \tag{31}$$

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. With that, there is no reason why the advisor 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 ($a \geq i^* \leftrightarrow L \geq 0$). 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:

$$\forall a \in S_{A;BR}: a \in [i^*, i_{max}]. \tag{32}$$

As both players know this, the investor can be expected to make an investment i equal 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 belief about the investor's mistrust respectively ($i_{guess} \leq a \leftrightarrow \bar{T}_{guess} \leq 0$). Based on this, the advisor can estimate their utility $U_A(a,i)$ by consulting their beliefs about the investor's type and making a guess i_{guess} on the investment i. This leads to the following function for the advisor's expected utility $U_A^e(a, i_{guess})$:

$$\begin{aligned} U_A^e(a, i_{guess}) &= U_A(a, i = i_{guess}) \\ &= \pi_A(i = i_{guess}) - C_L(a) \\ &= m_A * i_{guess} - C_L(a) \end{aligned} \tag{33}$$

with $m_A > 0$.

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

$$\bar{T}_{guess} = \frac{i_{guess} - a}{a} \quad \leftrightarrow \quad i_{guess} = (\bar{T}_{guess} + 1) * a \tag{34}$$

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

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

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

$$\begin{aligned} U_A^e\left(a, \bar{T}_{guess}(a)\right) &= m_A * i_{guess} - C_L(a) \\ &= m_A * \left(\bar{T}_{guess}(a) + 1\right) * a - C_L(a) \end{aligned} \tag{35}$$

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

Given this, the advisor can maximize their expected utility $U_A^e\left(a, \bar{T}_{guess}(a)\right)$ with respect to the advice $a \in [i^*, i_{max}]$. It yields:

$$\max_a \left(U_A^e\left(a, \bar{T}_{guess}(a)\right) \middle| \, a \in [i^*, i_{max}] \right) \quad \rightarrow \quad \frac{\partial U_A^e\left(a, \bar{T}_{guess}(a)\right)}{\partial a} \stackrel{!}{=} 0$$
$$\leftrightarrow \quad \frac{\partial \left(m_A * \left(\bar{T}_{guess}(a) + 1\right) * a - C_L(a)\right)}{\partial a} = 0 \qquad (36)$$
$$\leftrightarrow \quad \frac{\partial\left(C_L(a)\right)}{\partial a} = m_A * \frac{\partial \left(\left(\bar{T}_{guess}(a) + 1\right) * a\right)}{\partial a}$$

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

This means that any interior solution to the advisor's maximization problem must meet the condition that the derivative of the advisor's moral costs of lying $C_L(a)$ must be equal to the term $m_A * \frac{\partial\left(\left(\bar{T}_{guess}(a)+1\right)*a\right)}{\partial a}$. Therefore, any interior solution corresponds to giving advice a (between the optimal i^* and the maximal investment i_{max}) such that their 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). However, it is also possible that this maximization problem has a boundary solution (i.e., completely honest or completely dishonest behavior). For this reason, it could be that either giving completely honest advice a, equal to the optimal investment i^* ($a = i^*$), or giving the highest possible advice a, equal to the maximal investment i_{max} ($a = i_{max}$), solves this problem.

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

$$\forall a \in S_{A;BR}: \quad \frac{\partial\left(C_L(a)\right)}{\partial a} = m_A * \frac{\partial\left(\left(\bar{T}_{guess}(a)+1\right)*a\right)}{\partial a} \vee a = i^* \vee a = i_{max}. \qquad (37)$$

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

$$S_{A;BR} = \left\{ a^s \,\middle|\, a^s, a^{s\prime} \in \left\{ a \in [i^*, i_{max}] \,\middle|\, \frac{\partial\left(C_L(a)\right)}{\partial a} = m_A * \frac{\partial \left(\left(\bar{T}_{guess}(a) + 1\right) * a\right)}{\partial a} \vee a = i^* \vee a = i_{max} \right\} \wedge \forall a^{s\prime}: U_A^e(a^s) \geq U_A^e(a^{s\prime}) \right\} \qquad (38)$$

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

B.2. The investor's set of best responses $S_{I;BR}$

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

$$U_I(a,i) = \pi_I(i) - C_{\bar{T}}(a,i) = \begin{cases} i \leq i^*: & i_{max} + m_I * i - C_{\bar{T}}(a,i) \\ i > i^*: & i_{max} + m_I * (2 * i^* - i) - C_{\bar{T}}(a,i) \end{cases} \qquad (39)$$

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^* ($a \geq i^* \leftrightarrow L \geq 0$). As the investor anticipates this ($a \geq i^*_{guess} \leftrightarrow L_{guess} \geq 0$), he or she will always make an investment i less or equal to their received advice a ($i \leq a \leftrightarrow \bar{T} \leq 0$). In addition, there is no reason why the investor would make an investment i below their guess i^*_{guess} on the optimal investment i^*. As a consequence, the investor's set of best responses $S_{I;BR}$ consists only of investments i between their guessed optimal investment i^*_{guess} and the advice a. Hence:

$$\forall i \in S_{I;BR}: i \in \left[i^*_{guess}, a\right]. \tag{40}$$

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 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 $U_I^e\left(a,i,i^*_{guess}\right)$ as the following function of the advice a, their investment i, and their estimate of the optimal investment i^*_{guess}:

$$\begin{aligned} U_I^e\left(a,i,i^*_{guess}\right) &= \pi_I(i)_{i^*=i^*_{guess}} - C_{\overline{T}}(a,i) \\ &= \begin{cases} i \le i^*_{guess}: & i_{max} + m_I * i - C_{\overline{T}}(a,i) \\ i > i^*_{guess}: & i_{max} + m_I * \left(2 * i^*_{guess} - i\right) - C_{\overline{T}}(a,i) \end{cases} \end{aligned} \tag{41}$$

with $m_I > 0$.

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

$$L_{guess} = \frac{a - i^*_{guess}}{i^*_{guess}} \quad \leftrightarrow \quad i^*_{guess} = \frac{a}{L_{guess} + 1} \tag{42}$$

with $i^*_{guess} > 0$ and $L_{guess} \ge 0$.

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

$$\begin{aligned} U_I^e\left(a,i,L_{guess}\right) &= U_I^e\left(a,i,i^*_{guess} = \frac{a}{L_{guess}+1}\right) \\ &= \begin{cases} i \le \dfrac{a}{L_{guess}+1}: & i_{max} + m_I * i - C_{\overline{T}}(a,i) \\ i > \dfrac{a}{L_{guess}+1}: & i_{max} + m_I * \left(2 * \dfrac{a}{L_{guess}+1} - i\right) - C_{\overline{T}}(a,i) \end{cases} \\ &= \begin{cases} i < \dfrac{a}{L_{guess}+1}: & i_{max} + m_I * i - C_{\overline{T}}(a,i) \\ i \ge \dfrac{a}{L_{guess}+1}: & i_{max} + m_I * \left(2 * \dfrac{a}{L_{guess}+1} - i\right) - C_{\overline{T}}(a,i) \end{cases} \end{aligned} \tag{43}$$

with $m_I > 0$ and $L_{guess} \ge 0$.[98]

In order to maximize this function $U_I^e\left(a,i,L_{guess}\right)$ with regard to the investment i, it must be considered that this function's domain is split into two regions ($i < i^*_{guess} = \frac{a}{L_{guess}+1}$ and $i \ge i^*_{guess} = \frac{a}{L_{guess}+1}$). 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 $U_I^e\left(a,i,L_{guess}\right)$ 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 $C_{\overline{T}}(a,i)$ 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 $U_I^e\left(a,i,L_{guess}\right)$ within the second region ($i \ge i^*_{guess} = \frac{a}{L_{guess}+1}$).

[98] The last transformation of the expected utility $U_I^e\left(a,i,L_{guess}\right)$ is valid, since $U_I^e\left(a,i,L_{guess}\right)$ is a continuous function. For $i = i^*_{guess} = \frac{a}{L_{guess}+1}$ it yields: $i_{max} + m_I * i - C_{\overline{T}}(a,i) = i_{max} + m_I * \left(2 * \frac{a}{L_{guess}+1} - i\right) - C_{\overline{T}}(a,i)$ with $m_I > 0$ and $L_{guess} \ge 0$.

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:

$$\begin{aligned}\max_i\left(U_I^e(a, i, L_{guess}) \middle| i \in \left[\frac{a}{L_{guess}+1}, a\right]\right) \quad &\rightarrow \quad \frac{\partial U_I^e(a, i, L_{guess})}{\partial i} \overset{!}{=} 0 \\ &\leftrightarrow \quad -m_I - \frac{\partial\left(C_{\bar{T}}(a, i)\right)}{\partial i} = 0 \\ &\leftrightarrow \quad \frac{\partial\left(C_{\bar{T}}(a, i)\right)}{\partial i} = -m_I\end{aligned} \tag{44}$$

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

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

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

$$\forall i \in S_{I;BR}: \quad \frac{\partial\left(C_{\bar{T}}(a,i)\right)}{\partial i} = -m_I \vee i = \frac{a}{L_{guess}+1} \vee i = a. \tag{45}$$

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

$$\begin{aligned}S_{I;BR} = \Bigg\{ i^s \Bigg| i^s, i^{s\prime} \in \left\{ i \in \left[\frac{a}{L_{guess}+1}, a\right] \middle| \frac{\partial\left(C_{\bar{T}}(a, i)\right)}{\partial i} = -m_I \vee i = \frac{a}{L_{guess}+1} \vee i = a \right\} \\ \wedge \forall i^{s\prime}: U_I^e(i^s) \geq U_I^e(i^{s\prime}) \Bigg\}\end{aligned} \tag{46}$$

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

B.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 Continuous Deception Game is:

$$S_{equ} = \left\{ (a^s, i^s) \middle| \left(a^s, a^{s'} \in \left\{ a \in [i^*, i_{max}] \middle| \frac{\partial (C_L(a))}{\partial a} = m_A * \frac{\partial (i^s(a))}{\partial a} \vee a = i^* \vee a = i_{max} \right\} \wedge \forall a^{s'}: U_A^e(a^s) \geq U_A^e(a^{s'}) \right) \wedge \left(i^s, i^{s'} \in \left\{ i \in [i^*, a^s] \middle| \frac{\partial (C_{\bar{T}}(a^s, i))}{\partial i} = -m_I \vee i = i^* \vee i = a^s \right\} \wedge \forall i^{s'}: U_I^e(i^s) \geq U_I^e(i^{s'}) \right) \right\} \tag{47}$$

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

B.4. Further implications

So far, the cost functions ($C_L(a)$ and $C_{\bar{T}}(a,i)$) were defined very generally. To specify, 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:

$$\frac{\partial}{\partial a}\left(\frac{\partial (C_L(a))}{\partial a}\right) = \begin{cases} a < i^*: & \leq 0 \\ a > i^*: & \leq 0 \end{cases} \tag{48}$$

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

$$\frac{\partial}{\partial i}\left(\frac{\partial (C_{\bar{T}}(a,i))}{\partial i}\right) = \begin{cases} i < a: & \leq 0 \\ i > a: & \leq 0 \end{cases} \tag{49}$$

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 potential equilibria S_{equ} can be used to draw conclusions on the impact that each player's belief about the other ($\bar{T}_{guess}$ or, respectively, L_{guess}) has 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:

$$\frac{\partial (C_L(a))}{\partial a} = m_A * \frac{\partial (i(a))}{\partial a} \tag{50}$$

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 this game, 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:

$\lvert \bar{T}_{guess}(a) \rvert \uparrow$ (for all advice a)	$\rightarrow \quad \bar{T}_{guess}(a) \downarrow$	(since: $\bar{T}_{guess}(a) \leq 0$)
	$\rightarrow \quad i_{guess}(a) \downarrow$	(since: $i_{guess}(a) = \left(\bar{T}_{guess}(a) + 1\right) * a$)
	$\rightarrow \quad \dfrac{\partial\left(i_{guess}(a)\right)}{\partial a} \downarrow$	(since: $i_{guess}(a)$ becomes zero for $a = 0$, is concave, and increases monotonously.[99] Thus, if $\forall a$: $i_{guess}(a) \downarrow$, this function must be flatter. It follows that, for any increment of the advice a, the increment of $i_{guess}(a)$ must be lower.)
	$\rightarrow \quad \dfrac{\partial\left(i(a)\right)}{\partial a} \downarrow$	(since: $i_{guess} = i$ due to correct beliefs in equilibrium)
	$\rightarrow \quad \left(m_A * \dfrac{\partial\left(i(a)\right)}{\partial a}\right) \downarrow$	(since: $m_A > 0$)
	$\rightarrow \quad \dfrac{\partial\left(C_L(a)\right)}{\partial a} \downarrow$	(since: $\frac{\partial\left(C_L(a)\right)}{\partial a} = m_A * \frac{\partial\left(i(a)\right)}{\partial a}$ must be fulfilled for any interior solution to the advisor's maximization problem in equilibrium)
	$\rightarrow \quad C_L(a)$ accepted by advisor $\uparrow$	(since: $C_L(a)$ is concave and increases monotonously for $a \geq i^*$)
	$\rightarrow \quad a \uparrow$	(since: $C_L(a)$ increases monotonously for $a \geq i^*$)
	$\rightarrow \quad (a - i^*) \uparrow$	(since: i^* is constant)
	$\rightarrow \quad L \uparrow$	(since: $L \geq 0 \leftrightarrow a \geq i^*$).

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

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

$$\frac{\partial\left(C_{\bar{T}}(a,i)\right)}{\partial i} = -m_I \tag{51}$$

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:

[99] 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. Therefore, the advisor's guess i_{guess} on the investment i can be assumed to be a concave function of the advice a.

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

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

$$i^*_{guess} = \frac{a}{L_{guess} + 1} \tag{52}$$

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 are potentially considered by the investor in equilibrium, since:

$L_{guess} \uparrow$	$\rightarrow$	$\left(\frac{a}{L_{guess} + 1}\right) \downarrow$	(since: $a \geq 0$ and $L_{guess} \geq 0$)
	$\rightarrow$	$i^*_{guess} \downarrow$	(since: $i^*_{guess} = \frac{a}{L_{guess}+1}$)
	$\rightarrow$	$\min(i) \downarrow$	(since: $i \in \left[i^*_{guess}, a\right]$)
	$\rightarrow$	$\max(\lvert\bar{T}\rvert) \uparrow$	(since: $i \leq a \leftrightarrow \bar{T} \leq 0$ and $\bar{T} = \frac{i-a}{a}$).

In other words, in equilibrium a stronger suspicion of being lied to (expressed in L_{guess}) makes the investor consider more mistrusting strategies (i.e., a higher possible extent of risk-reducing mistrust $\lvert\bar{T}\rvert$ 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 C. Materials

C.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*.[100] Original German instructions are available upon request.

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

classEx

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 A1. Instructions for the advisor – Part 1/5.

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

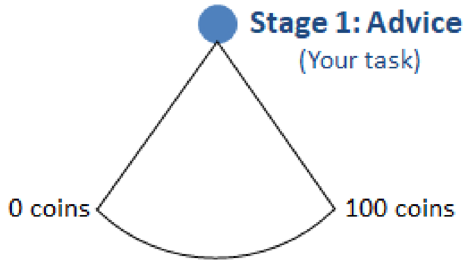
Procedure within a round:

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)

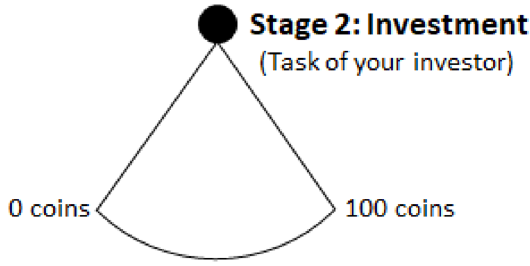


Figure A2. Instructions for the advisor – Part 2/5.

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.

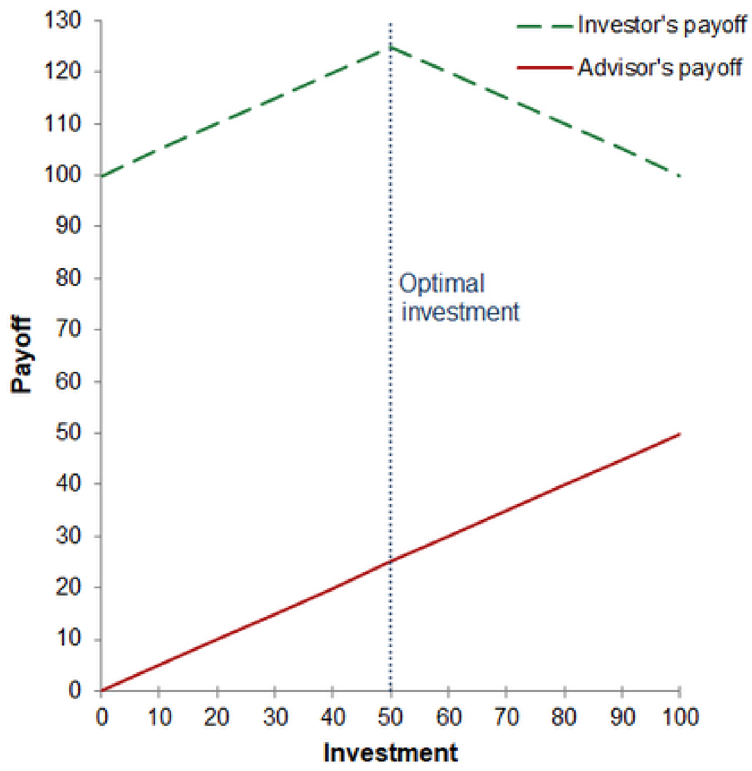
Optimal investment:

- **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 A3. 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 A4. 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 A5. Instructions for the advisor – Part 5/5.

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

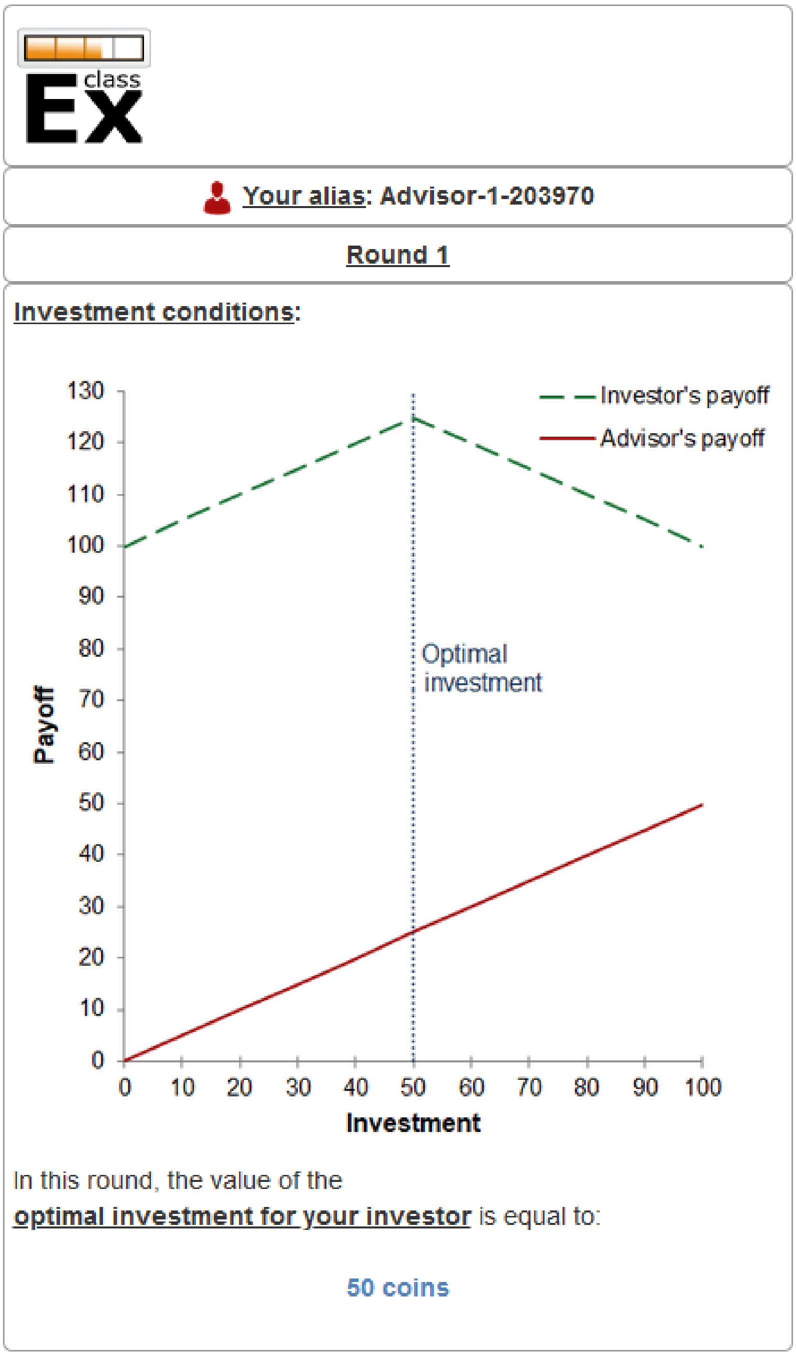


Figure A6. Input screen for the advisor – Part 1/2.

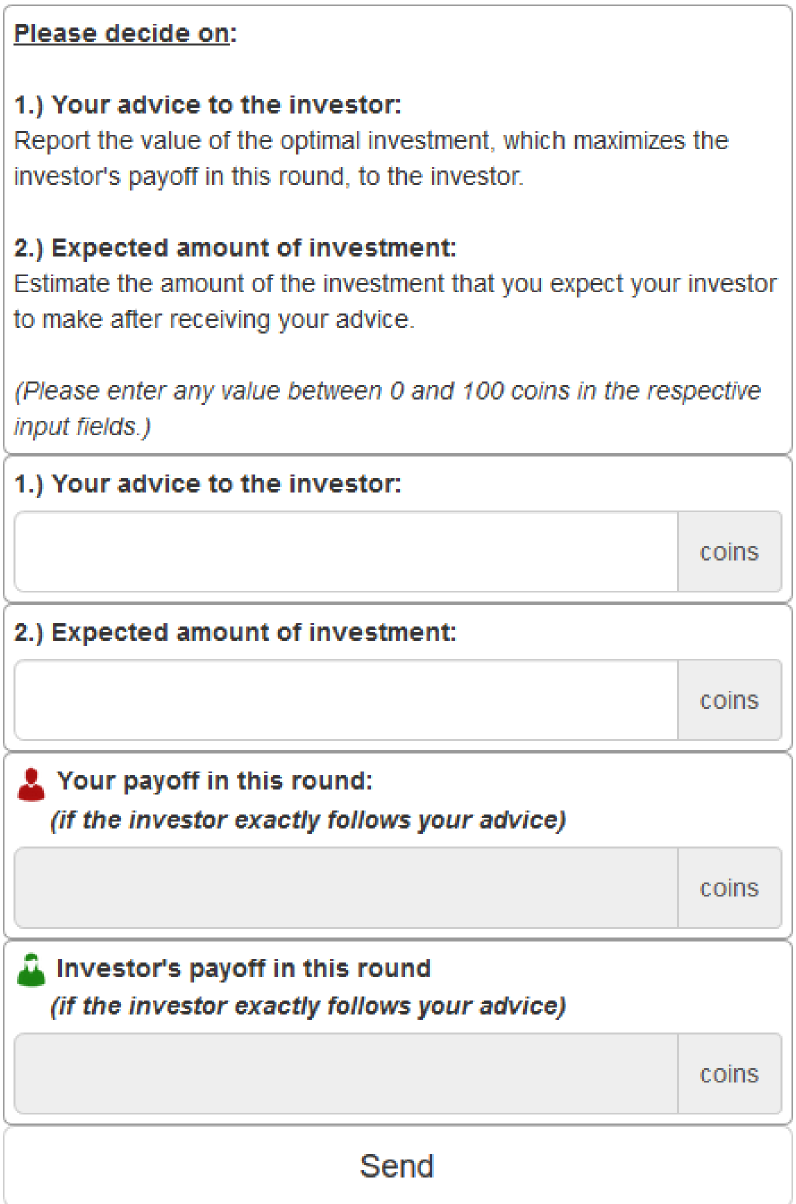


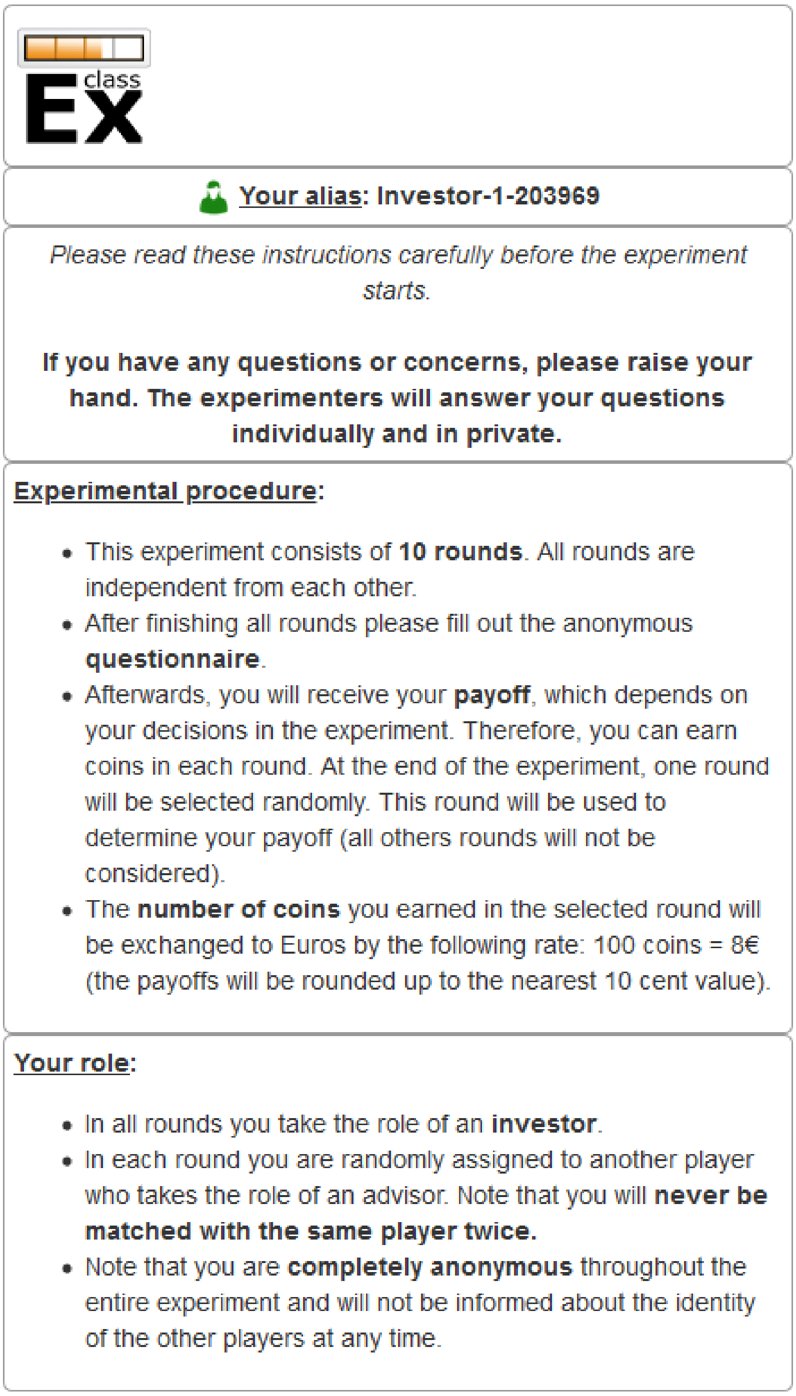
Figure A7. Input screen for the advisor – Part 2/2.

Note: The two greyed-out fields in Figure A7 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.

C.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*.[101] Original German instructions are available upon request.

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


class EX

Your alias: Investor-1-203969

Please read these instructions carefully before the experiment starts.

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

Experimental procedure:

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

Your role:

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

Figure A8. Instructions for the investor – Part 1/5.

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

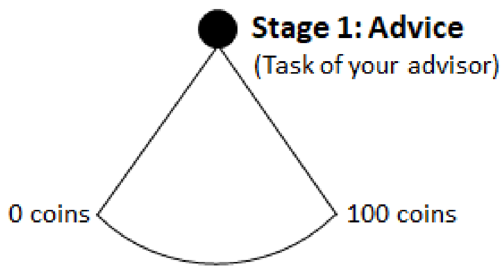
Procedure within a round:

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)

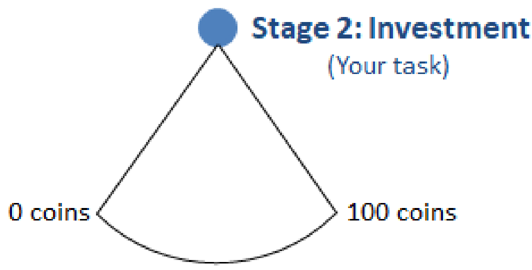


Figure A9. Instructions for the investor – Part 2/5.

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.

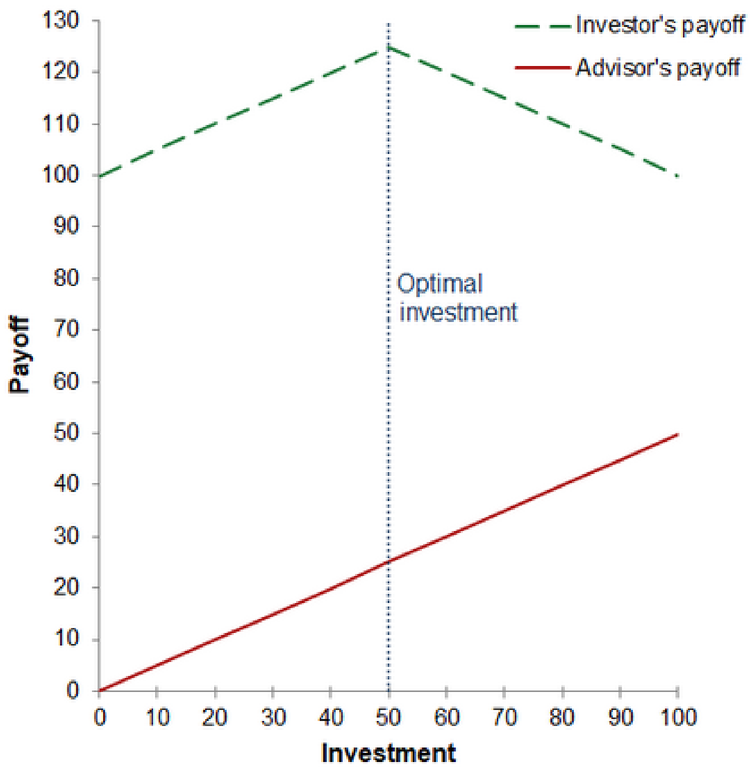
Optimal investment:

- **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 A10. Instructions for the investor – Part 3/5.

<u>**Investment conditions**</u>:

- **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 A11. 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 A12. Instructions for the investor – Part 5/5.

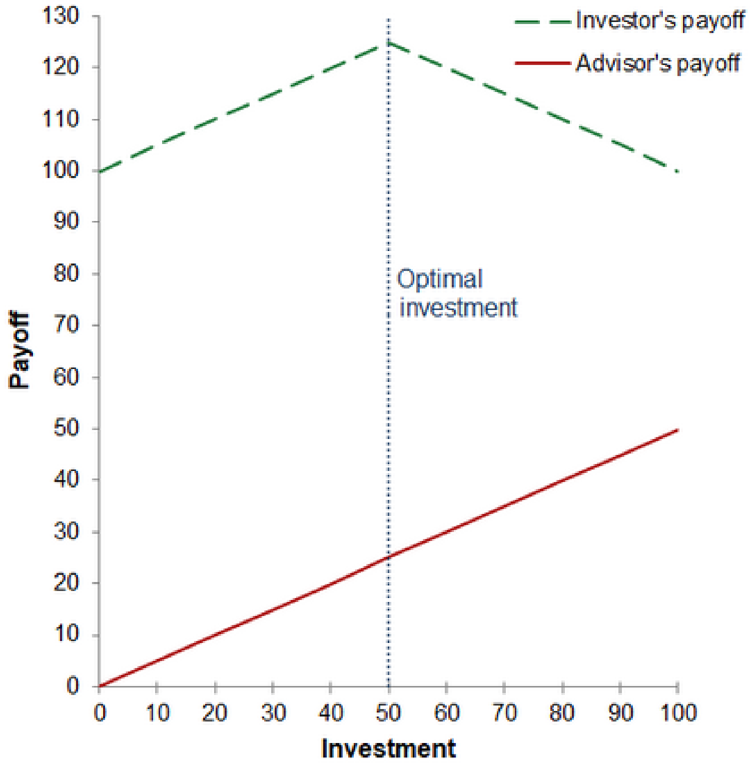
C.2.2. Input screen for one single round of the Continuous Deception Game (for the investor)

Your alias: Investor-1-203969

Round 1

Investment conditions:

In each round, the true value of the optimal investment for you is located somewhere between 0 and 100 coins.

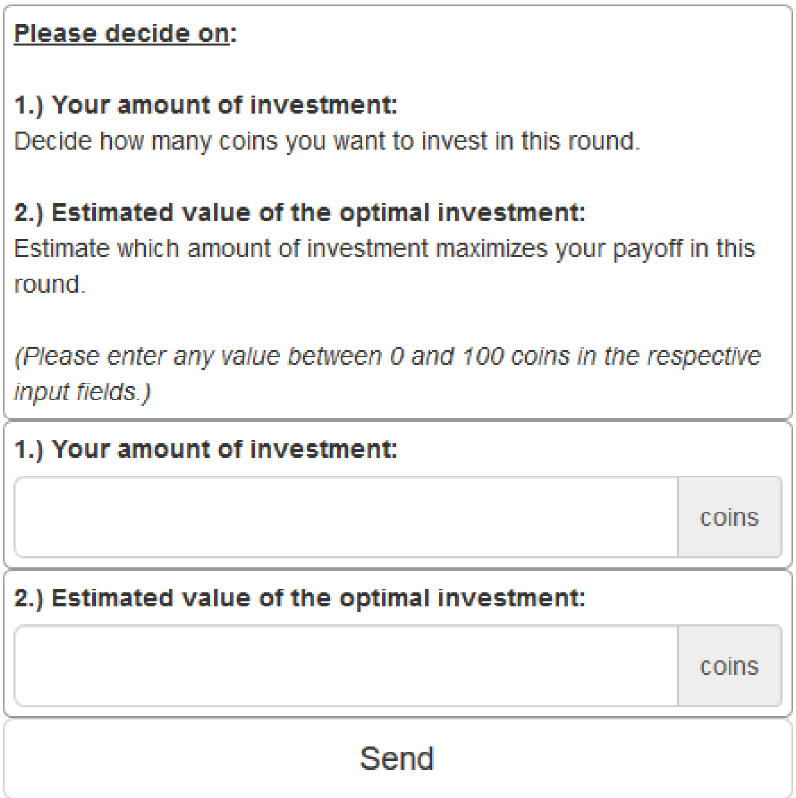


[Note to the reader, not to the subjects: In order to visualize that the value of the optimal investment could be located anywhere between 0 and 100 coins, I used an animated version of this figure in which the value of the optimal investment and the corresponding peak of the investor's payoff function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

Your advisor has reported that the **value of the optimal investment**, which maximizes your payoff in this round, is equal to:

50 coins

Figure A13. 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 A14. Input screen for the investor – Part 2/2.

References

1. Bank fined £10.5m over mis-selling. Availabe online: https://www.belfasttelegraph.co.uk/news/uk/bank-fined-10-5m-over-misselling-28688514.html (accessed on 8 April 2020).
2. SEC Charges Bernard L. Madoff for Multi-Billion Dollar Ponzi Scheme. Availabe online: https://www.sec.gov/news/press/2008/2008-293.htm (accessed on 8 April 2020).
3. SEC charges Massachusetts RIA with fraud. Availabe online: https://www.investmentnews.com/sec-charges-massachusetts-ria-with-fraud-80838 (accessed on 8 April 2020).
4. SEC Charges Investment Adviser With Fraud. Availabe online: https://www.sec.gov/news/press-release/2019-77 (accessed on 8 April 2020).
5. Dimmock, S.G.; Gerken, W.C. Predicting fraud by investment managers. *Journal of Financial Economics* **2012**, *105*, 153-173, doi:10.1016/j.jfineco.2012.01.002.
6. Gneezy, U. Deception: The Role of Consequences. *American Economic Review* **2005**, *95*, 384-394, doi:10.1257/0002828053828662.
7. Erat, S.; Gneezy, U. White Lies. *Management Science* **2012**, *58*, 723-733, doi:10.1287/mnsc.1110.1449.
8. Charness, G.; Dufwenberg, M. Promises and Partnership. *Econometrica* **2006**, *74*, 1579-1601, doi:10.1111/j.1468-0262.2006.00719.x.
9. Fischbacher, U.; Föllmi-Heusi, F. Lies in Disguise--An Experimental Study on Cheating. *Journal of the European Economic Association* **2013**, *11*, 525-547, doi:10.1111/jeea.12014.
10. Gneezy, U.; Kajackaite, A.; Sobel, J. Lying Aversion and the Size of the Lie. *American Economic Review* **2018**, *108*, 419-453, doi:10.1257/aer.20161553.
11. Hurkens, S.; Kartik, N. Would I lie to you? On social preferences and lying aversion. *Experimental Economics* **2009**, *12*, 180-192, doi:10.1007/s10683-008-9208-2.
12. López-Pérez, R.; Spiegelman, E. Why do people tell the truth? Experimental evidence for pure lie aversion. *Experimental Economics* **2013**, *16*, 233-247, doi:10.1007/s10683-012-9324-x.
13. Lundquist, T.; Ellingsen, T.; Gribbe, E.; Johannesson, M. The aversion to lying. *Journal of Economic Behavior & Organization* **2009**, *70*, 81-92, doi:10.1016/j.jebo.2009.02.010.
14. Mazar, N.; Amir, O.; Ariely, D. The Dishonesty of Honest People: A Theory of Self-Concept Maintenance. *Journal of Marketing Research* **2008**, *45*, 633-644, doi:10.1509/jmkr.45.6.633.
15. Vanberg, C. Why Do People Keep Their Promises? An Experimental Test of Two Explanations. *Econometrica* **2008**, *76*, 1467-1480, doi:10.3982/ecta7673.
16. Pruckner, G.J.; Sausgruber, R. Honesty on the Streets: A Field Study on Newspaper Purchasing. *Journal of the European Economic Association* **2013**, *11*, 661-679, doi:10.1111/jeea.12016.
17. Battigalli, P.; Charness, G.; Dufwenberg, M. Deception: The role of guilt. *Journal of Economic Behavior & Organization* **2013**, *93*, 227-232, doi:10.1016/j.jebo.2013.03.033.
18. Blok, V. The Power of Speech Acts: Reflections on a Performative Concept of Ethical Oaths in Economics and Business. *Review of Social Economy* **2013**, *71*, 187-208, doi:10.1080/00346764.2013.799965.
19. Alger, I.; Weibull, J.W. Homo Moralis--Preference Evolution Under Incomplete Information and Assortative Matching. *Econometrica* **2013**, *81*, 2269-2302, doi:10.3982/ecta10637.
20. Beck, T.; Bühren, C.; Frank, B.; Khachatryan, E. Can Honesty Oaths, Peer Interaction, or Monitoring Mitigate Lying? *Journal of Business Ethics* **2020**, *163*, 467-484, doi:10.1007/s10551-018-4030-z.
21. Beugelsdijk, S.; de Groot, H.L.F.; van Schaik, A.B.T.M. Trust and Economic Growth: A Robustness Analysis. *Oxford Economic Papers* **2004**, *56*, 118-134, doi:10.1093/oep/56.1.118.
22. Bjørnskov, C. How Does Social Trust Affect Economic Growth? *Southern Economic Journal* **2012**, *78*, 1346-1368, doi:10.4284/0038-4038-78.4.1346.
23. Knack, S.; Keefer, P. Does Social Capital Have an Economic Payoff? A Cross-Country Investigation. *The Quarterly Journal of Economics* **1997**, *112*, 1251-1288, doi:10.1162/003355300555475.
24. Boatright, J.R. Swearing to be Virtuous: The Prospects of a Banker's Oath. *Review of Social Economy* **2013**, *71*, 140-165, doi:10.1080/00346764.2013.800305.
25. Barefoot, J.C.; Maynard, K.E.; Beckham, J.C.; Brummett, B.H.; Hooker, K.; Siegler, I.C. Trust, Health, and Longevity. *Journal of Behavioral Medicine* **1998**, *21*, 517-526, doi:10.1023/a:1018792528008.

26. Rotter, J.B. A new scale for the measurement of interpersonal trust. *Journal of Personality* **1967**, *35*, 651-665, doi:10.1111/j.1467-6494.1967.tb01454.x.

27. Kuroki, M. Does Social Trust Increase Individual Happiness in Japan? *The Japanese Economic Review* **2011**, *62*, 444-459, doi:10.1111/j.1468-5876.2011.00533.x.

28. Gurtman, M.B. Trust, distrust, and interpersonal problems: A circumplex analysis. *Journal of Personality and Social Psychology* **1992**, *62*, 989-1002, doi:10.1037/0022-3514.62.6.989.

29. Sapienza, P.; Toldra-Simats, A.; Zingales, L. Understanding Trust. *The Economic Journal* **2013**, *123*, 1313-1332, doi:10.1111/ecoj.12036.

30. McKnight, H.D.; Chervany, N.L. Trust and Distrust Definitions: One Bite at a Time. In *Trust in Cyber-societies: Integrating the Human and Artificial Perspectives*, Falcone, R., Singh, M., Tan, Y.-H., Eds. Springer: Berlin/Heidelberg, Germany, 2001; Vol. 1, pp. 27-54.

31. Charness, G.; Dufwenberg, M. Bare promises: An experiment. *Economics Letters* **2010**, *107*, 281-283, doi:10.1016/j.econlet.2010.02.009.

32. Ismayilov, H.; Potters, J. Why do promises affect trustworthiness, or do they? *Experimental Economics* **2016**, *19*, 382-393, doi:10.1007/s10683-015-9444-1.

33. Servátka, M.; Tucker, S.; Vadovič, R. Building Trust--One Gift at a Time. *Games* **2011**, *2*, 412-433, doi:10.3390/g2040412.

34. Sobel, J. Signaling Games. In *Encyclopedia of Complexity and Systems Science*, Meyers, R.A., Ed. Springer: New York, USA, 2009; pp 8125-8139.

35. Leland, H.E.; Pyle, D.H. Informational Asymmetries, Financial Structure, and Financial Intermediation. *The Journal of Finance* **1977**, *32*, 371-387, doi:10.2307/2326770.

36. Armantier, O.; Boly, A. Can Corruption be Studied in the Lab? Comparing a Field and a Lab Experiment. *CIRANO – Scientific Publications* **2008**, *2008s-26*, doi:10.2139/ssrn.1324120.

37. Sánchez-Pagés, S.; Vorsatz, M. An experimental study of truth-telling in a sender-receiver game. *Games and Economic Behavior* **2007**, *61*, 86-112, doi:10.1016/j.geb.2006.10.014.

38. Dreber, A.; Johannesson, M. Gender differences in deception. *Economics Letters* **2008**, *99*, 197-199, doi:10.1016/j.econlet.2007.06.027.

39. Sutter, M. Deception Through Telling the Truth?! Experimental Evidence from Individuals and Teams. *The Economic Journal* **2009**, *119*, 47-60, doi:10.1111/j.1468-0297.2008.02205.x.

40. Peeters, R.; Vorsatz, M.; Walzl, M. Truth, Trust, and Sanctions: On Institutional Selection in Sender-Receiver Games. *The Scandinavian Journal of Economics* **2013**, *115*, 508-548, doi:10.1111/sjoe.12003.

41. Peeters, R.; Vorsatz, M.; Walzl, M. Beliefs and truth-telling: A laboratory experiment. *Journal of Economic Behavior & Organization* **2015**, *113*, 1-12, doi:10.1016/j.jebo.2015.02.009.

42. Jacquemet, N.; Luchini, S.; Shogren, J.F.; Zylbersztejn, A. Coordination with communication under oath. *Experimental Economics* **2018**, *21*, 627-649, doi:10.1007/s10683-016-9508-x.

43. Jacquemet, N.; Luchini, S.; Rosaz, J.; Shogren, J.F. Truth Telling Under Oath. *Management Science* **2019**, *65*, 426-438, doi:10.1287/mnsc.2017.2892.

44. Vranceanu, R.; Dubart, D. Deceitful communication in a sender-receiver experiment: Does everyone have a price? *Journal of Behavioral and Experimental Economics* **2019**, *79*, 43-52, doi:10.1016/j.socec.2019.01.005.

45. Gneezy, U.; Saccardo, S.; Serra-Garcia, M.; van Veldhuizen, R. Bribing the Self. *CESifo Working Papers* **2020**, *8065*.

46. Berg, J.; Dickhaut, J.; McCabe, K. Trust, Reciprocity, and Social History. *Games and Economic Behavior* **1995**, *10*, 122-142, doi:10.1006/game.1995.1027.

47. Camerer, C.F.; Hogarth, R.M. The Effects of Financial Incentives in Experiments: A Review and Capital-Labor-Production Framework. *Journal of Risk and Uncertainty* **1999**, *19*, 7-42, doi:10.1023/a:1007850605129.

48. Harsanyi, J.C. Games with Incomplete Information Played by "Bayesian" Players, I–III, Part I. The Basic Model. *Management Science* **1967**, *14*, 159-182, doi:10.1287/mnsc.14.3.159.

49. Harsanyi, J.C. Games with Incomplete Information Played by "Bayesian" Players, I–III, Part II. Bayesian Equilibrium Points. *Management Science* **1968**, *14*, 320-334, doi:10.1287/mnsc.14.5.320.

50. Harsanyi, J.C. Games with Incomplete Information Played by "Bayesian" Players, I–III, Part III. The Basic Probability Distribution of the Game. *Management Science* **1968**, *14*, 486-502, doi:10.1287/mnsc.14.7.486.

51. Engelmann, J.B.; Fehr, E. The slippery slope of dishonesty. *Nature Neuroscience* **2016**, *19*, 1543-1544, doi:10.1038/nn.4441.
52. Mörtenhuber, S.; Nicklisch, A.; Schnapp, K.-U. What Goes Around, Comes Around: Experimental Evidence on Exposed Lies. *Games* **2016**, *7*, 29, doi:10.3390/g7040029.
53. Festinger, L. *A Theory of Cognitive Dissonance*; Stanford University Press: Stanford, CA, 1957.
54. Batson, C.D.; Thompson, E.R.; Seuferling, G.; Whitney, H.; Strongman, J.A. Moral hypocrisy: Appearing moral to oneself without being so. *Journal of Personality and Social Psychology* **1999**, *77*, 525-537, doi:10.1037/0022-3514.77.3.525.
55. Utikal, V.; Fischbacher, U. Disadvantageous lies in individual decisions. *Journal of Economic Behavior & Organization* **2013**, *85*, 108-111, doi:10.1016/j.jebo.2012.11.011.
56. Fisher, R.A. Theory of Statistical Estimation. *Mathematical Proceedings of the Cambridge Philosophical Society* **1925**, *22*, 700-725, doi:10.1017/s0305004100009580.
57. Vanberg, C. Who never tells a lie? *Experimental Economics* **2017**, *20*, 448-459, doi:10.1007/s10683-016-9491-2.
58. Cox, J.C. How to identify trust and reciprocity. *Games and Economic Behavior* **2004**, *46*, 260-281, doi:10.1016/s0899-8256(03)00119-2.
59. Innocenti, A.; Pazienza, M.G. Altruism and Gender in the Trust Game. *Labsi Working Papers* **2006**, *5/2006*.