

Lohse, Tim; Qari, Salmai

Working Paper

Gender Differences in Face-to-Face Deceptive Behavior

CESifo Working Paper, No. 7995

Provided in Cooperation with:

Ifo Institute – Leibniz Institute for Economic Research at the University of Munich

Suggested Citation: Lohse, Tim; Qari, Salmai (2019) : Gender Differences in Face-to-Face Deceptive Behavior, CESifo Working Paper, No. 7995, Center for Economic Studies and Ifo Institute (CESifo), Munich

This Version is available at:

<https://hdl.handle.net/10419/214997>

Standard-Nutzungsbedingungen:

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

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

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

Terms of use:

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

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

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

Gender Differences in Face-to-Face Deceptive Behavior

Tim Lohse, Salmal Qari

Impressum:

CESifo Working Papers

ISSN 2364-1428 (electronic version)

Publisher and distributor: Munich Society for the Promotion of Economic Research - CESifo GmbH

The international platform of Ludwigs-Maximilians University's Center for Economic Studies and the ifo Institute

Poschingerstr. 5, 81679 Munich, Germany

Telephone +49 (0)89 2180-2740, Telefax +49 (0)89 2180-17845, email office@cesifo.de

Editor: Clemens Fuest

www.cesifo-group.org/wp

An electronic version of the paper may be downloaded

- from the SSRN website: www.SSRN.com
- from the RePEc website: www.RePEc.org
- from the CESifo website: www.CESifo-group.org/wp

Gender Differences in Face-to-Face Deceptive Behavior

Abstract

We study the role of face-to-face interaction for gender differences in deceptive behavior and perceived honesty. In the first part, we compare women's to men's deceptive behavior using data from an incentivized income reporting experiment in which lies can be detected in the course of an audit. Between the three treatments of that experiment, (i) the degree, and (ii) the impact of the face-to-face interaction vary from none (computerized baseline treatment) to a little (treatment in which face-to-face communication triggers psychological effects such as greater lying aversion) to much (treatment in which the perception by others also enters as a strategic effect determining the probability of detecting a lie). In the computerized baseline treatment men and women lie alike. Women's truthfulness increases when psychological effects of face-to-face interaction come into play. In contrast, male deceptive behavior does not change until the strategic effect of perceived honesty matters and men's truthfulness rises way beyond the level of women. To elaborate on these gender differences, in the second part, participants are asked to assess the honesty of videotaped statements from an experimental setting identical to the third treatment. We find that more men are assessed as rather dishonest. Men's dishonest perception is independent of whether they are actually truthful or not and whether they are assessed by men or women. We conclude that men anticipate their low perceived honesty in a face-to-face setting and, therefore, deceive less compared to women.

JEL-Codes: C910, D910, J160.

Keywords: gender differences, lying, face-to-face interaction, honesty assessment, perception, video analysis, laboratory experiment.

Tim Lohse
Berlin School of Economics and Law
Badensche Str. 52
Germany – 10825 Berlin
tim.lohse@hwr-berlin.de

Salmal Qari
Berlin School of Economics and Law
Badensche Str. 52
Germany – 10825 Berlin
salmal.qari@hwr-berlin.de

December 4, 2019

This work was generously supported by the Joachim Herz Stiftung. Tim Lohse is grateful for travel funding by the Fritz Thyssen Stiftung.

1 Introduction

In many situations of political, business or private life, lying can be a tempting strategy. A candidate running for a political office might promise to bring back long-lost jobs. The CEO of a carmaker could praise the environmental compatibility of its actually manipulated diesel engines, or a financial adviser might suggest an allegedly safe investment knowing that the only sure thing would be her commission. In these examples, deceptive interaction can occur by telling fairytales on Twitter, publishing delusive technical reports to calm environmental protection agencies, or emailing beautified prospectuses to naive investors, respectively. Alternatively, deception opportunities may also arise in face-to-face settings. When canvassing, a politician might impress voters at their front door, some top-level managers could be particularly convincing in personal discussions with the regulator, and advisers could charm supposedly critical clients in-person. This paper compares deceptive behavior in settings with and without face-to-face interaction. Since women are known to be rather responsive to changes in the forms of communication (Conrads and Lotz, 2015) and men generally exhibit a higher degree of overconfidence that might play a crucial role when interacting in-person (Barber and Odean, 2001, among others), we are focusing on gender differences in deceptive behavior.

Behavioral differences in lying between situations with and without face-to-face interaction can be attributed to at least two channels. First, an individual's social image is particularly at stake when considering lying face-to-face. This can, for instance, cause a greater lying aversion due to higher costs of lying (Lundquist et al., 2009; Abeler et al., 2014; Kajackaite and Gneezy, 2017), increased guilt aversion (Charness and Dufwenberg, 2006; Vanberg, 2008) or invoke shame (Coricelli et al., 2014). We refer to these aspects as the *psychological effects* of face-to-face interaction. As men and women may feel differently about these effects, differences in deceptive behavior can be gender-specific. Second, deceiving individuals might also strive to be perceived as honest because the detection of a lie could imply severe consequences. In the above-mentioned examples, a liar runs the risk of losing an election, of high fines that might put the entire firm in danger or of being sentenced to jail, respectively. Chances of getting caught lying are most likely different in situations without face-to-face interaction compared to those with. In the latter, the appearance and performance of an individual matter. They determine

whether the individual is perceived as dishonest, which may entail an audit and, hence, a detection of the lie.¹ Thus, the possibility to influence one’s own lie detection probability can alter an individual’s deceptive behavior, too. We refer to this second channel as the *strategic effect* of perceived honesty. Differences in honesty perception on the one hand and in actual deceptive behavior on the other are essentially two sides of the same coin. If there are systematic or gender-specific differences in perception then there will be room for strategic behavior. Whoever is perceived as rather honest, and recognizes the way they are perceived, can take advantage and deceive with little chance of getting caught. These considerations are at the heart of our paper.

We study two laboratory experiments. In the first experiment, we compare women’s to men’s deceptive behavior using income reporting data. Between the three treatments of that experiment, we vary on the one hand the presence of the psychological effects and on the other hand the presence of the strategic effect of face-to-face interaction: income reporting is fully computerized in a baseline treatment but occurs face-to-face in the second and third treatment. However, the strategic effect of an individual’s perceived honesty impacting its individual audit probability is only present in the latter treatment. The research questions to be answered by such a setting are: does deceptive behavior depend on whether interaction occurs face-to-face? Are there gender differences? Are potential gender differences in deception induced by the psychological effects of face-to-face interaction or by the strategic effect of perceived honesty?

Our results are as follows. In the computerized baseline treatment women’s and men’s rate of honest reports is statistically indistinguishable (0.501 and 0.429, respectively). When the psychological effects are present, only women’s truthfulness increases significantly to 0.647. When additionally the strategic effect of an individual’s perceived honesty plays a role, women’s honesty slightly deteriorates to 0.57, whereas men’s truthfulness rises sharply to 0.769. Taken together, women behave significantly more truthfully when switching from an anonymous setting to one where only the psychological effects of face-to-face interaction matter. In contrast, men behave significantly more truthfully when the perceived honesty has an additional strategic effect.

The second experiment elaborates on what could drive these gender differences in deceptive

¹Key clues that allow for detecting deception in written messages are identified by Chen and Houser (2017).

behavior. In particular we investigate how women compared to men are perceived when lying face-to-face and, vice versa, how women compared to men are perceived when telling the truth. Moreover, we analyze whether women (men) perceive men (women) differently from how they perceive women (men). A gender-balanced set of participants watched videotaped statements from an experimental setting identical to the third treatment of the first experiment. They had to assess the honesty of the recorded female and male subjects. Evaluating more than 9,000 assessments reveals that more men are assessed as dishonest, both by women and men. Interestingly, men assess women as significantly less honest than women who assess other women. Moreover, men also appear more dishonest than women no matter whether they make a truthful statement or not.

These findings answer our research questions. Face-to-face interaction leads to more honest behavior but alters women's and men's deceptive behavior through different channels. Women are sensitive to the mode of communication and react to the experimental implementation of psychological effects of face-to-face interaction by becoming more truthful. In contrast, men do not change their behavior to that treatment variation, but strongly increase their truthfulness once the strategic effect of perceived honesty plays a role. Our finding of women decreasing their deceptive behavior drastically as soon as interaction occurs in-person indicates that they are concerned about their social image. In contrast, men do not seem to care too much about their social image. But they seem to shy away from a 'social competition' on appearing most honestly once psychological as well as strategic effects are present. As the assessment experiment shows, they do well by abstaining from lying in the face-to-face situation since they always appear less honest than women. Their individual probability of an audit and, thus, their chance of getting caught lying, is much higher than for women. We suspect that men correctly anticipate their low perceived honesty in a face-to-face setting and, therefore, deceive less whereas women take advantage of being perceived as honest and deceive more than men.

The remainder of the paper is organized as follows: first, we review some related papers (section 2) and explain the set-up of our experiments (section 3). Section 4 presents the results of the experiments and section 5 discusses the relation to other fields of study. Finally, we conclude (section 6).

2 Related literature

Our analysis contributes to the rapidly growing literature on dishonest behavior in general (Jacobsen et al., 2018, for a recent overview), and on behavioral dissimilarities between women and men in particular. Some field experiments reveal that women exhibit weaker cheating intentions in exams (Tibbetts, 1999), are more likely to return excess change (Azar et al., 2013; Bersoff, 1999), and are less likely to conduct fare dodging (Buccioli et al., 2013). Evidence from the laboratory points in the same direction of women acting more honestly than men (Houser et al., 2012; Kocher et al., 2018). Psychological explanations for such a gender gap in misconduct are women's higher degree of self-control leading them to disobey rules less frequently (Burton Jr et al., 1998; Higgins, 2004), men's learning of positive attitudes toward offending when being exposed to delinquent peers during adolescence (Mears et al., 1998; Akers, 2017), and women's internalization of negative emotions compared to men's externalization through delinquencies (Broidy and Agnew, 1997). In addition, women's higher social value orientation (Grosch and Rau, 2017), or their lower immoral intentions due to gender differences in emotion (Ward and King, 2018) can play a role. Therefore, women have a stronger tendency to feel shame from actions that deviate from honesty (Rebellion et al., 2015), particularly when the moral cost is high (Fosgaard et al., 2013). Also parents' impulse for honest behavior affects girls and boys differently (Buccioli and Piovesan, 2011; Houser et al., 2016).

However, various studies raise doubts about women's higher truthfulness. Some studies do not find evidence of gender differences in deceptive behavior at all (Childs, 2012; Ezquerra et al., 2018; Pate, 2018). Others find mixed evidence and discuss the potential impact of the gender composition of groups (Muehlheusser et al., 2015) or the implications of who benefits from a lie (Biziou-Van-Pol et al., 2015; Cappelen et al., 2013; Capraro, 2018; Dreber and Johannesson, 2008; Erat and Gneezy, 2012).

Our main innovation to the literature is the careful introduction of face-to-face interaction as a further dimension that could be decisive for gender differences in deceptive behavior. Usually anonymous and computerized settings are employed. However, lowering the social distance by interacting in-person may, e.g., increase prosocial behavior (Charness et al., 2007; Charness and Gneezy, 2008). In fact, recent research suggests that deception occurs less frequently when

the degree of anonymity between parties is lowered due to audio communication (Cohn et al., 2018). Reducing anonymity even further via face-to-face interaction does increase honesty, but, as our findings show, it works for men and women through two different channels. So far, only a few experimental deception papers have allowed for face-to-face interaction but without analyzing the gender dimension. They look at the impact of temporally distancing the decision task from the payment (Ruffle and Tobol, 2014), subjects' reluctance to provide accurate negative feedback (Gneezy et al., 2017), the ability to detect deceit (Belot and Van De Ven, 2017; Holm and Kawagoe, 2010; Serra-Garcia and Gneezy, 2019), the consequences of endogenizing the audit probability (Konrad et al., 2017), or the probability of success in deluding others (Dwenger and Lohse, 2019). Our insights highlight that the environment to study gender differences in deception crucially matters.

3 The experimental set-up

Our paper consists of two experiments: first, we study gender differences in deceptive behavior using data from an incentivized income reporting experiment with three treatments. We refer to this experiment as the social interaction and honesty experiment since the degree of social interaction and its implications for potential audits vary between the treatments.² Second, we run an honesty-assessment experiment in which a gender-balanced set of participants was asked to assess the honesty of videotaped statements of men and women. The link between both parts is that the experimental setting in which the videos were recorded was identical to one of the treatments of the social interaction and honesty experiment. In the following, we describe each of the two parts in more detail. The two parts of our experimental study were run at two different university laboratories on two distinct subject pools.

²We use the data from our accompanying paper Konrad et al. (2017). In that paper, we focused on the consequences of endogenizing the audit probability in a face-to-face setting but did not look into potential gender differences in reporting behavior and their underlying effects, which is the focus of this study.

3.1 The social interaction and honesty experiment

3.1.1 *The income reporting game with room for deception*

Subjects play a simple income reporting game involving a binary choice about whether to lie or not. At the beginning of each round, individuals sit in front of a computer and learn about their randomly assigned private income. The value displayed on the monitor is private information and it is either high (1,000 experimental currency units (ECU), with probability 0.8) or low (ECU 400, with probability 0.2). Individuals have to self-report an income to the (tax) authorities. Authorities cannot observe the individual's income directly, but know the probability distribution. The reporting individual can choose between two possible statements: high or low income. If an individual reports a high income, a tax of ECU 200 is due. Individuals with low income have a strictly dominant strategy not to cheat, but to report low income. This report resulted in a zero statutory tax liability. In contrast, an individual with a high income has a true choice: either to honestly state the high income or to deceive by reporting a low income. Deception leads to two possible outcomes: On the one hand, the individual might be selected for an audit. If an audit occurs, it is carried out by the computer and perfectly reveals the individual's true income. In this case, as the lie is detected, the individual has to pay taxes of ECU 200 topped by a fine. In half of all sessions the fine was ECU 100 (which makes lies worthwhile in expectation for a risk-neutral subject), and in the other half it was ECU 300.³ On the other hand, the individual might not be detected because no audit occurred. Then the subject pays neither taxes nor a fine. Throughout all treatments, an individual gets to know at the end of each round whether an audit occurred or not. Successful lies occur at the expense of the experimenter and can therefore be classified as black lies.

3.1.2 *The experimental treatments*

Applying a between-subjects design, there are three different treatments that we label baseline (BL), social interaction (SI), and social interaction & honesty (SIH). The treatments differ in how the income report is made – with or without face-to-face social interaction between an individual and tax authorities – and how the detection mechanism works – with or without an

³A higher fine leads to a higher level of compliance, but this is not the focus of the present paper.

individual's perceived honesty impacting its audit probability. The honesty-assessment experiment explained in the next subsection is based on videos that were recorded in a treatment identical to SIH.

In the fully computerized baseline treatment (BL) there is no social interaction. The individual is asked on the computer screen to report either a high or a low income. The high-income individual knows (as this is written in the instructions) that the computer randomly detects deceivers with probability 0.5. The subjects make their reporting decisions in a laboratory room in which 20 subjects perform the same task independently. As the detection probability of 0.5 is fixed, each subject's task is formally independent of the tasks and choices of other individuals. The social interaction treatment (SI) allows face-to-face interaction between subjects and a person in the role of a (tax) officer. The 20 subjects, mixed females and males, again first learn their income while sitting in front of a computer in the same laboratory room as in BL. They were then called up individually in a message on their screen and were sent into a neighboring room to state their report. The sequence as well as the gender composition in which the subjects were called up was determined randomly. There were two reporting rooms, allowing for two subjects at a time to be called up for a private interview. In the reporting room an officer greeted the subject, identified them by their identification number (preserving pseudonymity throughout), and then asked for an oral report of either a low or a high income. The officer entered the report into the computer system and the subject returned to the laboratory. Thus, the conversation between an officer and a reporting subject was rather short and standardized which allowed us to focus on gender differences in deceptive behavior rather than e.g., differences in creativity for making up long stories.⁴ The first round is over once 10 subjects have reported to one officer and the other 10 subjects to the other officer. In each round, the officer is a man in one room and a woman in the other room. Subjects meet each of the four officers only once. The audit mechanism is the same as in BL: in SI the high-income subject also knows (as this is written in the instructions) that the probability of being detected when lying is 0.5 with the computer system making a random selection. The officer has no active decision role and does not learn whether an individual will be audited or

⁴Belot and Van De Ven (2017) show that longer communication (90 seconds instead of only 10 seconds) does not improve deception detection.

not. The SI treatment takes into account that subjects who report face-to-face to a real person rather than to a computer as in BL may behave differently. Differences between BL and SI can be attributed to the social interaction and the related psychological effects of the face-to-face reporting decision explained above. The SI treatment controls for the sum of these psychological effects and facilitates distinguishing them from the strategic effect of perceived honesty which comes into play in the third treatment (SIH). The social interaction & honesty treatment (SIH) is similar in structure to the SI treatment. The crucial difference is that detection does not occur randomly as in BL and SI. Participants knew *ex ante* that the officers influenced the choice of whom to select for an audit: officers had to assess a series of up to 10 reports and grade subjects with respect to perceived honesty (grades range from 1 = very honest to 10 = very dishonest). Then, in a first step, the computer system identified subjects with a high income but reported a low one. This identification was not communicated to anyone (neither to the officers nor to the reporting subject or other participants) and was purely a technical necessity for the second step. In that second step, half of the deceiving subjects were selected for an audit. This selection occurred according to the ranking made by the officer: the computer system selected those 50% of underreporting subjects for an audit who were perceived by the officer as being the most dishonest. High-income subjects who considered deceiving thus had to take into account how their appearance and lying performance would affect their individual audit probability, knowing that half of all deceivers would get detected. Subjects who deem themselves as rather dishonest might do best by reporting truthfully whereas subjects who believe they are perceived as honest in general have a (monetary) incentive to take advantage of their positive perception and deceive with a low risk of getting caught.⁵ As in the two other treatments, at the end of each round subjects learn about their payoff and, thereby, whether a potential lie was detected in the course of an audit or not. Officers do neither get feedback about their grading nor do they learn who was audited. This third treatment captures the strategic effect of perceived honesty. It is of particular interest to see whether this treatment variation affects men and women differently. A possible behavioral prediction could be that men deceive more. The reason could be that they may be overconfident about their honest

⁵Note that the audit mechanism ensures that the aggregate detection rate (0.5) is constant across the three treatments for all potential liars and is the same as in BL and SI, thereby removing a potential confounder of the treatment effects.

appearance and, thus, believe that their individual audit probability is comparably low.

3.1.3 Procedure of the experiment

The 12 sessions of the experiment – four for each treatment – were conducted at the Munich Experimental Laboratory for Economic and Social Sciences (MELESSA). Participants were recruited using the software ORSEE (Greiner, 2015). In each session there were 20 subjects, totaling a number of 240. The participating subjects were students of diverse fields at the University of Munich. Since more women than men are registered as potential participants in the MELESSA’s subject pool, more women participate in all treatments (BL: 51 vs. 29; SI: 46 vs. 34; SIH: 49 vs. 31).

Before the experiment started, instructions were read aloud and handed out in print. The instructions, which were identical for all participants of a session, explained the experimental setting in detail (see Appendix). After reading the instructions, all participants had to complete a quiz in order to make sure they fully understood the experimental set-up. Each individual participated in only one of the three treatments. Participants in the computerized BL treatment played the income reporting game for 10 independent rounds. Since each round in the two treatments with social interaction, SI and SIH, takes more time than a round in BL, participants in SI and in SIH played exactly four independent rounds.⁶ Independence was induced by a replacement of the person who served as an officer between rounds. Recall, that each participant met each of the four officers only once. After the experiment, the subjects had to answer an exit questionnaire on their gender and their field of study, etc. They were also asked to self-evaluate their perceived honesty and, thus, state their beliefs about their individual audit probability in the SIH treatment. This was followed by an incentivized standard risk elicitation game in the style of Holt and Laury (2002). Finally, participants received their earnings from the experiment (one randomly selected round), plus the outcome of the risk elicitation game, plus a show-up fee of EUR 4. Average earning was EUR 16.59 (standard

⁶In the treatment BL, ECU 1,000 were converted into EUR 10. In SI and SIH, ECU 1,000 were converted into EUR 16. These different exchange rates ensured that the participants’ expected payoffs per unit of time were the same if they showed the same choice behavior in all treatments. Section 4.1.3 also presents an analysis using only the data from the first round of the treatments. This analysis corroborates the main results. Further, Drichoutis et al. (2015) provide evidence that the exact conversion rate hardly matters for experimental behavior.

deviation 3.59), with men earning EUR 16.41 and women EUR 16.71 on average, respectively.

To ensure anonymity for the subjects, the four officers (two men and two women) in SI and SIH, respectively, were not recruited from the pool of student subjects. Instead, young employees and contract workers from the Max Planck Institute for Tax Law and Public Finance played the role of tax officers. In line with reality, they were paid a flat fee. Moreover, since the research focus is entirely on the subjects' reporting behavior, a further incentivization of the officers was not necessary. The participating subjects were neither informed about the officers' payment schemes nor that they were amateurs in their role.

3.2 The honesty-assessment experiment

In the honesty-assessment experiment we invited a large number of participants to assess the gender-specific honesty of some videotaped statements recorded in an experimental setting similar to the SIH treatment. The gender balanced set of invited assessors rated the video clips as 'rather honest' or 'rather dishonest', which allowed us to generate a general honesty score for each videotaped subject as we describe below.

3.2.1 Generating and selecting the videos

We ran the exact experimental setting of the SIH treatment and videotaped (including voice recordings) the subjects' statements as part of a larger research project (Dwenger and Lohse, 2019; Konrad, Lohse and Qari, 2014; Lohse and Qari, 2018). The videos were generated in several waves at the MELESSA. Recruitment happened in the same way as for the social interaction and honesty experiment except that people who had previously participated in that experiment were not invited. Again participants were almost exclusively students from diverse fields of study. A clip took about 20 seconds. Since each participant played four rounds, there are four video clips per reporting subject. In total, we generated 472 videoclips.⁷ Depending on the randomly assigned income, a subject in SIH is potentially videotaped in three different

⁷Participants were informed about the recording before signing up. They had to give their written consent that the videotapes could be used for scientific purposes while strict pseudonymity would be preserved. They were free to revoke their consent and to quit the experiment at any time. But no one did. Privacy is also kept in the honesty-assessment experiment and there is no feedback provided on whether a specific clip shows an honest or a dishonest statement.

situations: honestly reporting high income, honestly reporting low income, and dishonestly reporting low income when in fact having a high income. The two latter types of video clips are particularly useful for our honesty assessment experiment.

For this paper, we use the maximal subset of participants given the following requirements. Each participant should be featured only once. Moreover, the number of truthful women and truthful men as well as the number of untruthful women and untruthful men, respectively, should be the same. According to these requirements, we randomly selected video clips among those that were taped multiple times and created a unique and perfectly balanced set of 80 clips featuring 80 subjects (with an average age of about 23 years). This set shows 40 women and 40 men who all make the same statement (i.e., they report a low income). However, for each gender 20 videos show deceiving subjects, and 20 videos show truthful reports, respectively. This approach is based on the finding that the use of video-camera recording does neither affect the composition of the set of participants nor does it impact the observed decision behavior that we are interested in (Lohse and Qari, 2018).

We divide the set of 80 clips into two disjoint subsets of 48 (subset 1) and 32 clips (subset 2), respectively.⁸ Each subset is perfectly balanced with respect to the gender composition and (dis)honesty of statements.

3.2.2 *The outcome measure*

For each of the video clips shown we generate an honesty score (HS) as

$$HS = \frac{\# \text{ of times the video clip is assessed as "rather honest"}}{\text{total } \# \text{ of assessments}}. \quad (1)$$

The honesty score can range from 0 to 1: A value of 1 indicates that all assessors evaluate the videotaped statement as being honest. In contrast, if all of them assess the recorded statement as dishonest the value will be 0. This score allows us to answer the following questions: Are the score distributions of videotaped women and men different? Does the score distribution of videotaped women depend on the gender of the assessor, and, equivalently,

⁸This split happens along the dimension of the fine in the SIH treatment which is of no particular interest to our honesty-assessment experiment. In subset 1 containing 48 clips the fine was low, and in subset 2 it was high.

does the distribution of videotaped men depend on the assessors gender? In particular, we will analyze whether features like the minimum score, the maximum score, the average and the variance are gender-dependent.

3.2.3 Procedure of the experiment

The honesty-assessment experiment took place at the experimental laboratory of the Technical University (TU) Berlin. By recruiting participants for the videotaped SIH treatment in Munich and those for the honesty-assessment experiment in Berlin, we made sure that (i) there was no overlap of subjects, and (ii) the assessors did not personally know the videotaped individuals. Subjects were recruited from the student subject pool of the TU's experimental laboratory in a similar way as in Munich using the software ORSEE (Greiner, 2015). 232 subjects took part in the experiment, grouped in 10 sessions with up to 24 participants. The two subsets of video clips were shown to 116 assessors each. Each session consisted equally of male and female assessors. This gives a total of 9,280 assessments.

At the beginning of the assessment experiment, detailed instructions were read aloud and handed out in print (see Appendix). The instructions explained the task of assessing the video clips, the sequential but random order of video clips, and the fact that each clip could only be assessed once. Each assessor sits in front of a computer screen and has a headphone to listen to the subject's income statement. After having watched the video clip the assessor is asked to make a binary assessment, i.e., to assess whether (s)he perceives a subject's stated report of a low income as honest or not. Assessors have to pass judgment on the clip displayed before moving on to the next video clip. There is no possibility to return and change the assessment.⁹ Finally, participants have to fill in an exit questionnaire.

At the end of the experiment, five out of all assessments were randomly selected to determine the assessor's payoff. Each video clip correctly assessed was worth EUR 5 for subset 1 and EUR 4 for subset 2, respectively. That is, the assessment task could earn participants up to EUR 25 or EUR 20, respectively, plus a fixed show-up fee of EUR 5. On average, participants

⁹To minimize heterogeneity among the assessors they also explicated that the ratio of honest to dishonest reports in the set of video clips shown was about 50/50 (thus, among assessors the median fraction of videos rated as rather untruthful is equal to 0.52).

earned EUR 16.83 (EUR 17.41 for subset 1, and EUR 16.24 for subset 2) with a standard deviation of EUR 5.08 (EUR 5.69 and EUR 4.35, respectively), including the show-up fee.

4 Results

4.1 Results of Experiment I: The social interaction and honesty experiment

4.1.1 Descriptives

We first consider descriptives of the main outcome variable, i.e. we calculate the fraction of honest reports separately for women and men. For women, this fraction is equal to 0.50 in BL. It is equal to 0.66 in SI and, respectively, 0.57 in SIH. Looking at men, the fraction of truthful reports is equal to 0.41 in BL. Compared to this baseline the percentage of truthfully reporting men is slightly larger in SI (0.49) and much higher in SIH (0.77). Hence, the descriptives clearly suggest that men change their behavior only slightly between BL and SI, while men strongly react to the strategic component –the possible assessment of their honesty– induced by treatment SIH. The descriptives for women suggest a treatment effect of about 16 percentage points between BL and SI and a treatment effect of about –9 percentage points between SI and SIH. These numbers yield two observations: First, women distinguish between the mode of reporting, i.e., anonymous reporting in BL vs. face-to-face social interaction in SI and SIH. Second, the small treatment effect between SI and SIH suggests that women react slightly to the possible assessment of their honesty by increasing deception. In summary, the descriptives suggest that the effect of perceived honesty induced by treatment SIH causes men to strongly reduce their deceptive behavior while women slightly increase their level of deception.

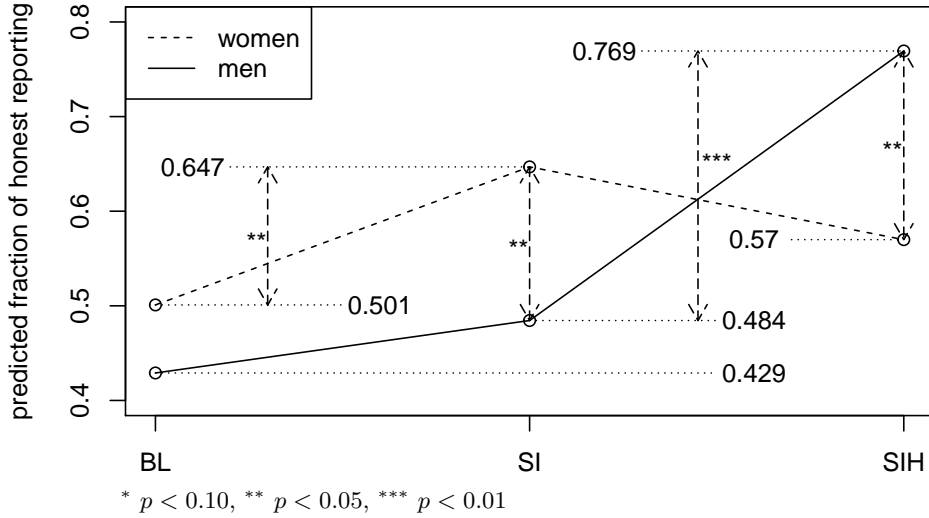
4.1.2 Regression Results

We now discuss the results from a random-effects logit-model. The dependent variable is the dummy indicating whether the subject is making a truthful statement. Formally, the equation reads

$$y_{it} = \Xi_i' \beta + u_i + \epsilon_{it}. \quad (2)$$

The explanatory variables collected in Ξ_i are dummy variables indicating female subjects, the three treatments and interaction terms between the treatment indicators and the female dummy. u_i denotes subject-specific random-effects that are included to control for the dependency of observations for each subject in the course of the several rounds. The logit-model predicts the probability of reporting truthful as follows: $P(Y_{it} = 1|u_i, \Xi_i) = \Lambda(\Xi_i' \beta + u_i)$ where $\Lambda(\cdot)$ denotes the logistic cumulative distribution function.

Fig 1: Predicted effects – fraction of honest reports by gender



Column (1) of table 1 presents the main regression results, and Figure 1 summarizes the associated predicted effects. Overall, the results corroborate the descriptives. The intercept (-0.167) models the probability/fraction of truthfully reporting men in SI. Evaluating this intercept yields that the predicted fraction of truthfully reporting men is equal to 0.484 in SI.¹⁰ The insignificant *BL*-coefficient indicates that men’s behavior in BL is statistically not distinguishable from their behavior in SI. The *SIH*-coefficient models the huge observed shift for men in SIH; the predicted fraction of honest men in SIH is equal to 0.769. The *Female*-coefficient shows that women in SI have a significantly larger probability (0.647) of reporting truthfully compared to men in SI. The two remaining interaction terms model the additional shift in the probability of reporting truthfully for women in BL and SIH. Evaluating the

¹⁰We report average marginal effects (i.e., population-averaged effects) that are comparable to the descriptive statistics. We obtain the effects numerically by first simulating 500,000 draws of the subject-specific-random effects and then averaging over the 500,000 subject-specific-marginal effects. For example, a single subject-specific probability of reporting truthfully for a male subject in SI is $\Lambda(-0.167 + u_i)$ where $\Lambda(\cdot)$ denotes the logistic cumulative distribution function.

Table 1: Probability of reporting honestly – random effects logit model

	(1)	(2)	(3)	(4)
	Basis spec	+Riskmeasure	+Age	+Penalty
BL	-0.587 (0.874)	-0.266 (0.876)	-0.619 (0.882)	-0.352 (0.822)
SIH	3.258*** (0.959)	3.474*** (0.970)	3.333*** (0.966)	2.963*** (0.923)
Female	1.750** (0.843)	1.740** (0.837)	1.580* (0.836)	1.654** (0.797)
BL * Female	-0.985 (1.145)	-1.119 (1.140)	-0.811 (1.138)	-1.008 (1.061)
SIH * Female	-4.101*** (1.248)	-4.152*** (1.245)	-4.023*** (1.238)	-3.512*** (1.174)
RiskMeasure		0.304** (0.132)	0.275** (0.132)	0.260** (0.122)
Age ∈ [21, 23]			-0.528 (0.774)	-0.556 (0.722)
Age ∈ [24, 27]			0.571 (0.802)	0.169 (0.749)
Age ∈ [28, 31]			1.080 (1.035)	0.328 (0.968)
Age ≥ 32			0.707 (1.311)	-0.493 (1.252)
High Penalty				3.266*** (0.504)
Constant	-0.167 (0.611)	-0.339 (0.613)	-0.271 (0.892)	-1.780** (0.877)
σ_u^2	14.47*** (3.725)	14.03*** (3.586)	13.49*** (3.427)	9.283*** (2.260)
N	1155	1155	1155	1155
Log pseudo lik.	-564.4	-562.6	-560.6	-538.9
Chi-squared	20.44	23.76	28.01	56.66
prob>Chi2	0.00103	0.000579	0.00180	3.84e-08

Notes: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$, standard errors in parentheses. This table presents the results from a random effects logit model that predicts the individual probability of honestly reporting a high income. The predictors are dummies indicating the treatments (BL and SIH; SI is the omitted reference category), a dummy indicating female subjects and interaction terms. Further controls are a measure for risk aversion and dummies indicating age (age ≤ 20 is the omitted reference category) and a dummy indicating a high penalty for underreporting. Random effects on the subject-level are included to take into account the repeated observations for the same subjects.

Summary: The results show that men do not change their reporting behavior between treatment SI (the omitted baseline category) and treatment BL, i.e., men do not react to the psychological effects of face-to-face communication. Men’s truthfulness increases strongly in treatment SIH, i.e., men react strongly to the strategic effect of perceived honesty. Men in treatment SIH a significantly more truthful than women in treatment SIH. Women’s truthfulness is similar in treatments SI and SIH. Furthermore, women’s truthfulness in SI is larger than in BL. This suggests that women react to the mode of communication (face-to-face vs. anonymous communication), but do not react to the strategic effects of perceived honesty. Specifications (2) to (4) show that the results of the baseline specification (1) are robust.

respective sums of coefficients yields that the (predicted) percentage of truthfully reporting women is equal to 0.501 in BL, 0.647 in SI, and 0.57 in SIH. While the difference between 0.647 and 0.57 (SIH and SI) is not significant, the associated p-value for the difference between BL and SI is equal to 0.036.¹¹ Furthermore, the percentage of truthfully reporting women in SIH (0.57) is significantly smaller than the corresponding percentage for men (0.769). Hence, as in the descriptives, women adjust their behavior in response to the mode of declaration (SI compared to BL). Unlike men, women’s behavior in SIH is very similar to SI, suggesting that women do not react to the effect of perceived honesty induced by SIH. We summarize our results so far as follows:

Result 1. *When introducing face-to-face interaction, but only psychological effects matter, it is just women’s truthfulness, which increases. When, additionally, the strategic effect of an individual’s perceived honesty plays a role, it is men’s truthfulness that rises sharply and exceeds women’s level by far.*

4.1.3 Robustness Checks

We further inquire as to the robustness of the results obtained so far. First, we run a number of regressions entering background characteristics such as the age of a subject or a measure of risk aversion. Columns (2) to (4) of table 1 present these results. In summary, all additional specifications confirm our two main findings. First, women’ reporting behavior is similar in both treatments involving social interaction (SI and SIH) suggesting that women react to the mode of reporting but barely react to the perceived honesty competition component of the SIH treatment. Second, men strongly react to the strategic effect of perceived honesty in treatment SIH by reducing their deceptive behavior. Second, we check whether our results are affected by behavioral adjustments over time. To this end, we analyze the reporting behavior in the first round of the experiment. Figure S1 in the online appendix indicates that we obtain the same results: men do not change reporting behavior between BL and SI, but strongly between SI and SIH. Running a logit model shows that this difference is once again statistically significant (see Table S2 in the online appendix). Further, this logit model indicates that women adjust

¹¹Table S1 in the supplementary appendix provides the regression results that allow us to obtain this p-value.

their behavior significantly between SI and BL, but not between SIH and SI.¹²

4.1.4 Self-Assessment

The observed gender differences raise the question of why men react so strongly to the effect of perceived honesty. A first answer stems from participants of the SIH treatment who were asked to assess their own subjective probability of being audited. They had to state whether they believed that their subjective probability of receiving an audit was smaller, equal to or larger than 50 percent. About 47 percent of women and 42 percent of men state that their individual audit probability is smaller than 50 percent. Moreover, while 41 percent of women believe that their audit probability is equal to 50 percent, this applies to only 32 percent of men. Finally, the answers indicate that the share of men who believe that their audit probability is larger than 50 percent is twice as large as the corresponding share of women (26 percent of men vs. 12 percent of women). These numbers suggest that men compared to women are more likely to believe that they ‘lose the social competition’ of being perceived as honest. This finding motivates our second experiment.

4.2 Results of Experiment II: The honesty-assessment experiment

4.2.1 Descriptives

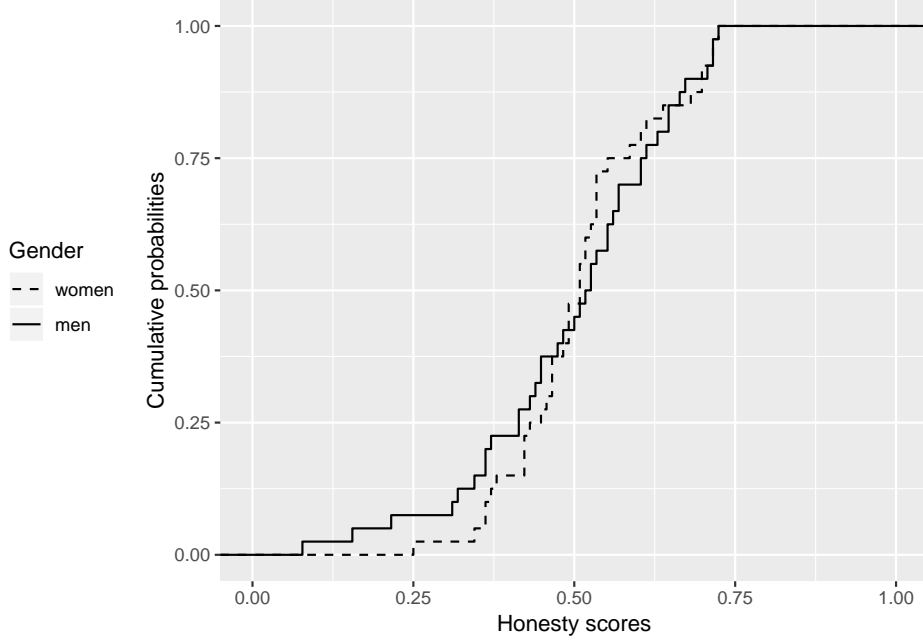
As explained in section 3.2, the subjects of this second experiment were asked to assess the honesty of the videotaped statements. In total, there are 80 statements from subjects who were videotaped while participating in a treatment similar to SIH from the first experiment, and each video was assessed by 116 assessors. We first compute the honesty score (HS) for each of the 80 videos according to equation (1).

Figure 2 provides the (empirical) cumulative distribution functions of the honesty scores. There are two immediate findings: First, the distribution of men’s honesty scores has a larger probability mass for very small honesty values. For example, the sample minimum for men’s honesty scores is equal to 0.0776, while the corresponding minimum for women is equal to 0.25. This suggests that the least honest-looking woman is still perceived as 3.5 times more honest

¹²Table S2 and S3 in the online appendix provide the results. The two tables present the same logit model, but use a different references categories that allow to test directly the respective differences.

than the most dishonest-looking man. Second, the largest honesty score is virtually identical for women and men and equal to 0.724. In turn, this suggests that the dispersion of honesty scores is larger for men compared to women.

Fig 2: Cumulative distribution functions of honesty scores by gender



Summary: The figure indicates that the average scores for women and men are similar (close to 0.5) and that the variance and the lower tails of the distributions are different. For example, it shows that the probability mass for women’s honesty score is equal to zero for values below 0.25, i.e., the smallest honesty score for women is equal to 0.25. The corresponding smallest honesty score for men is equal to 0.0776. This indicates that some men but no women are assessed as dishonest by virtually all judging subjects.

A further related descriptive observation concerns potential stochastic dominance of the two distributions. As the two cumulative distribution functions cross each other a few times, there is no indication of first-order stochastic dominance. Since the average honesty score for women is larger than the average for men, we checked second-order stochastic dominance (SOSD). Let $G_f(hs) = \text{Prob}(HS \leq hs)$ denote the cumulative distribution function (cdf) of women’s honesty scores, and let $G_m(\cdot)$ denote the corresponding cdf for men’s honesty scores. We compute the area under $G_f(hs)$ and, respectively, under $G_m(hs)$ for each value of hs and find that the area under $G_f(hs)$ never exceeds the area under $G_m(hs)$. More formally, these

calculations imply that

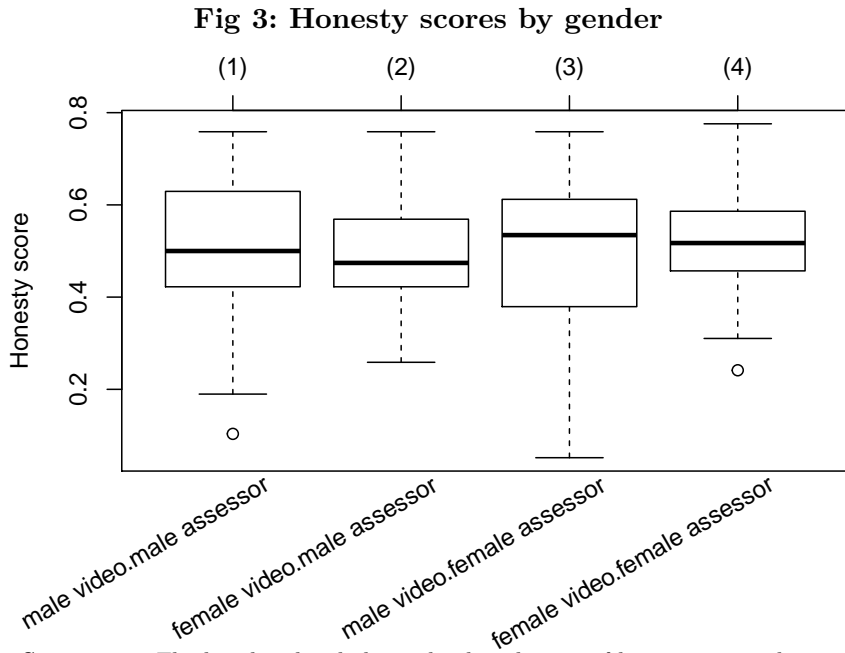
$$\int_{-\infty}^{hs} G_m(y) - G_f(y) dy \geq 0 \quad (3)$$

holds for each hs and with strict inequality for at least one hs which constitutes a further descriptive observation: The distribution of women’s honesty scores second-order stochastically dominates the distribution of men’s honesty scores.

Taken together, the two descriptive findings suggest that men face a higher likelihood of being evaluated as rather dishonest. In turn, this provides a first piece of evidence that could explain why men reduce their deceptive behavior so strongly in treatment SIH of experiment I.

4.2.2 Assessment by assessor’s gender

We now exploit the balanced structure of our dataset and probe whether the assessments of the videotaped women and men depend on the assessor’s gender. In particular, we analyze whether the variance and the mean of the honesty score distributions is different for female and, respectively, male assessors.



Summary: The boxplots break down the distribution of honesty scores by gender of the videotaped subject and gender of the assessing subject. Boxes 1 and 3 indicate that the men’s minimum honesty score is rather low, i.e., only (some) men are assessed as dishonest by almost all assessors. Consequently, the boxplots suggest that the variance of honesty scores is larger for men. Further, the medians suggest that women benefit from being assessed by women rather than by men (box 4 vs box 2).

Figure 3 depicts the distribution of honesty scores broken down by the gender of the videotaped subject and assessor. Boxes 1 and 3 provide the scores of videotaped men, while boxes 2 and 4 show the distribution of videotaped women. The overall comparison of the four boxplots is in line with the evidence so far: the maximum honesty score is similar for women and men and independent of the assessor’s gender. The minimum honesty score for men is clearly smaller than the corresponding minimum score for women. Statistical tests concerning the equality of variances indicate that the variance differs between videotaped men and women (p-value of 0.044), but does not depend on the assessors’ gender. Therefore, the tests confirm the visual findings.

The boxplots further suggest small differences between the median honesty scores. For example, comparing boxes 1 and 3 suggests that the median honesty score of videotaped men is slightly larger if men are assessed by a woman rather than by a man. However, the comparison further suggests that women assess men more frequently as rather dishonest: the minimum honesty score in box 3 is clearly smaller than the one in box 1 and the minimum in box 3 is not considered as an outlier. Comparing boxes 2 and 4 yields a similar pattern: the median honesty score of videotaped women is slightly larger if women are assessed by a woman. Further, box 4 indicates a small shift compared to box 2: the 50 percent of the observations at the center of the distribution are shifted toward the top and therefore indicate higher honesty scores, i.e., women are perceived as more honest by female assessors compared to male assessors. We test whether there are statistically significant differences between the average honesty scores of the four groups by running a logit model that predicts the probability that a videotaped subject i will be assessed as honest by assessor j . Formally, the equation reads

$$v_{ij} = \alpha_0 + \alpha_1 * FemaleVideo_i + \alpha_2 * FemaleAssessor_j + \alpha_3 * FemaleVideo * FemaleAssessor_{ij} + \epsilon_{ij} \quad (4)$$

where v_{ij} is equal to one if the subject is assessed as honest by assessor j , $FemaleVideo$ is a dummy variable indicating whether or not subject i is female, $FemaleAssessor$ is a dummy indicating whether or not assessor j is female and $FemaleVideo * FemaleAssessor$ is the

interaction term. Recall that the descriptive analysis so far suggests a fair amount of correlation regarding the assessments (e.g., the minimum honesty score for men is 0.0776). We employ robust standard errors clustered on the level of videotaped subjects to take this into account. The results (see table S4 in the supplementary appendix) indicate that for videotaped women there is a significant difference between the assessment by women rather than men (p-value 0.033), i.e., videotaped women are assessed as more honest by women compared to men. These results complement the findings obtained from the boxplots, and we summarize our findings as follows:

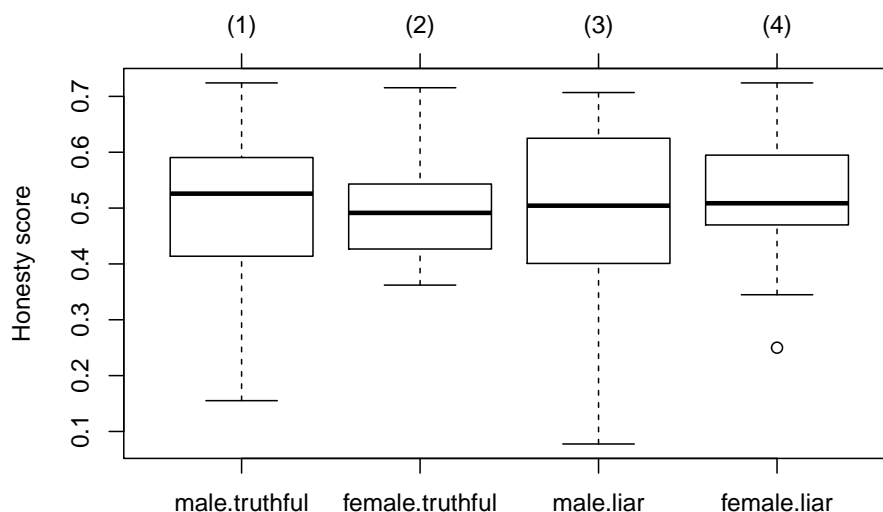
Result 2. *The variance of male honesty scores is larger than the variance of female honesty scores. Thus, men face the possibility of being assessed as dishonest by virtually all assessors; this possibility does not depend on the assessors' gender. Moreover, men's predicted average honesty score is unrelated to the assessors' gender. In contrast, women receive a higher predicted honesty score if they are assessed by a woman rather than a man.*

4.2.3 Assessment by deceptive behavior

Recall that our experimental set-up allows us to verify ex-post whether or not a videotaped subject lied. Figure 4 breaks down the distribution of honesty scores by gender and deceptive behavior. This allows us to analyze –for example– whether or not the minimum honesty score is different for subjects who report truthfully compared to subjects who lie. The boxplots suggest that the distribution does not strongly depend on deceptive behavior. In line with the empirical distribution functions (Figure 2), the dispersion of the male distribution is larger compared to the female distribution. Furthermore, both the minimum and the maximum honesty score are similar for truthful and lying subjects. Once again, we use a regression-based test to assess the differences between the variances. The test indicates that the observed differences between the variances are statistically significant (the one-sided p-value is 0.043.). We summarize these findings as follows:

Result 3. *The distribution of honesty scores is unrelated to subjects' actual deceptive behavior. Men face a higher probability of being perceived as rather dishonest and this does not depend on their deceptive behavior.*

Fig 4: Honesty scores by deceptive behavior



Summary: The boxplots break down the distribution of honesty scores by gender and the actual deceptive behavior of the videotaped subjects. The plots indicate that the variance of honesty scores is larger for men than for women. Once again, the larger variance for men stems from a much smaller minimum honesty score compared to women. Furthermore, the distribution of honesty scores is very similar for truthful and lying subjects.

5 Discussion

A major reason why face-to-face interaction impacts deceptive behavior is undoubtedly the reduction of social distance. Comparing honesty in chat and audio communication, subjects' truthfulness increases as anonymity decreases (Cohn et al., 2018). Our analysis builds on this finding and takes it a decisive step further to shed light on gender differences in deceptive behavior. Face-to-face interaction essentially minimizes the social distance, and our experimental set-up allows us to disentangle the related psychological and strategic aspects. This is new to the literature which has recently shown theoretically that an aversion to lying might hinge not only on an intrinsic cost of lying but also on social image concerns (Abeler et al., 2019; Dufwenberg and Dufwenberg, 2018; Gneezy et al., 2018; Khalmetski and Sliwka, 2019). Such reputational concerns are certainly more present when subjects interact face-to-face. From a naive point of view the observed overall increase in honesty from the fully computerized baseline treatment to the SIH treatment does not seem to come as a surprise. However, it would be a misjudgment to ignore the two kinds of aspects that work rather differently for

men and women. Our results of the SI treatment reveal that it is only women that respond to the reduction in social distance and, hence, show concerns about their reputation. The fact that men do not significantly vary their deceptive behavior between baseline and SI could be interpreted as experimental evidence that social image concerns seem to be more alien to them. To elaborate on these differences even further, of course, a variety of intermediate treatments between BL and SI would have been conceivable such as a real-time chat, first anonymously and, second, with a face picture of the counterpart, or a real-time video chat. We are unable to describe the potentially associated even finer behavioral changes and can only study the aggregate change.

A somewhat open question is why women take advantage of their less dishonest perception and deceive more than men in the SIH treatment. One possible explanation could be that women have been socially disadvantaged for centuries (Di Cagno et al., 2016). Persons belonging to such groups that historically have been discriminated against are less likely to trust and, therefore, might behave more strategically (Alesina and La Ferrara, 2002). These are far-reaching considerations which might provide an explanation for the observed gender differences in the respective treatment. However, these considerations are to some extent at odds with the existing evidence that women exhibit a higher social value orientation (Grosch and Rau, 2017), and such an explanatory approach is beyond the limits of our experimental design, which prevents us from shedding more light on this topic.

Finally, we are able to link our findings to the experimental literature on gender differences in competitiveness (cf. the review by Niederle, 2016). For example, in the well-known matrix task –on average– women perform worse than men in a competitive environment where the performance is monitored by the experimenter (Gino et al., 2013). In a setting where individuals can choose whether or not their performance is monitored, more women than men opt for less monitoring and rather misreport their performance. The authors interpret their findings as evidence that women anticipate their disadvantage and therefore avoid (more often than men) the competitive environment. However, the competitiveness literature remains surprisingly silent about competitive behavior in the actual workplace which is widely characterized by face-to-face interaction. Many situations of business life can be considered a competition with face-to-face interaction whose outcome is determined by how one individual, in comparison to

other individuals, is perceived by a third party. To make the point clear, think of a manager who wants to contract an external consultant. All applying consultants are eager to be hired. So they present themselves one by one trying to convince the manager that (s)he would be the best. The consultants can decide to lie, for example, by neglecting risks or overstating potential gains because they know that the manager hardly has time to check the proposed ideas and concepts. Rather, their perceived honesty is vital for the manager's decision. In a nutshell, one could refer to such a situation as a "social competition to appear most honest" and our results suggest that there are substantial gender differences. Women seem to be sensitive to the introduction of face-to-face interaction per se and change their behavior whereas men are more likely to be concerned about the introduction of the 'social competition'. In fact, they appear to shy away from competing for honesty by behaving more truthfully in SIH as they are correctly afraid of losing the social competition. This is fairly at odds with the conventional wisdom of men exhibiting greater overconfidence and women being averse to competition. The reason for our rather different result is that the standard insights usually draw upon experiments whose participants had to perform a variety of tasks related to one's own sporting or intellectual abilities. However, as pointed out by Croson and Gneezy "it is hard to know how sensitive the results are to the task used" (Croson and Gneezy (2009) p. 465). In fact, our findings urge caution regarding the experimental tasks used to investigate gender differences. Face-to-face interaction seems to be crucial and may potentially even reverse standard results.

6 Conclusion

This paper investigates experimentally how male and female deceptive behavior differs in situations with face-to-face interaction and what causes such gender differences. Allowing face-to-face interaction potentially impacts deception via two channels: first, psychological effects such as an aversion to lying might be more pronounced in a face-to-face situation. Second, the opportunity to influence one's audit probability by one's appearance and performance represents a strategic effect. In line with previous literature, women and men lie alike in the absence of face-to-face interaction. When introducing such interaction and controlling for the first channel, women lie less, which provides evidence of women's concern about their social

image. In contrast, as men do not change their deceptive behavior, it seems that they do not care too much about their social image per se. When additionally allowing for the second channel where a subject's perceived honesty plays a role for potentially detecting a lie, men's truthfulness rises sharply. In fact, when both channels are present, women deceive significantly more than men do. We argue that this could be due to gender differences in perceived honesty by others. Both genders assess more men as less honest. The likelihood of being perceived as rather dishonest is larger for men than for women and is independent of whether they are actually truthful or not. Hence, in a face-to-face reporting setting, the individual probability of an audit, and thus the chance of getting caught lying, is higher for men. Men might anticipate their low perceived honesty in a face-to-face setting and, therefore, deceive less compared to women.

Given recent compelling evidence concerning the external validity of laboratory experiments on cheating (Dai et al., 2018), our findings do not only speak to theorists but should also be whispered in the ears of princes (in the words of Roth, 2015). Voters facing candidates in town hall meetings, members of Congress hearing testimony from a carmaker's executive, or investors talking to their financial adviser at the counter should resist their guts about whom they perceived as especially honest. In situations with face-to-face interaction, perception and actual behavior can easily fall apart. Hence, by picking those for a scrutiny whom are deemed as rather dishonest voters, regulators or investors might falsely focus on the honest guys and just miss the truly bad ones. In particular, our findings raise concern that there could be profound gender differences. An audit strategy relying on random selections seems to be more promising.

Acknowledgments We thank Uri Gneezy, Luisa Herbst, Charles Holt, Sebastian Kessing, James Konow, Kai A. Konrad, Ulrich Schmidt, and Marta Serra-Garcia, seminar participants at the Kiel Institute for the World Economy, the Rady School of Management, University of California San Diego, the Loyola Marymount University in Los Angeles, the University of Siegen, the University of Salzburg and conference participants of the annual meeting of the Southern Economic Association in Washington, DC, and the European ESA meeting in Dijon for very helpful comments and fruitful discussions. For providing laboratory resources

we kindly thank the TU-Lab of the Technical University of Berlin and the MELESSA of the University of Munich. We also thank Hans Mueller for developing and programming the web-based environment. Tim Lohse is grateful for hospitality at the Rady School of Management, University of California San Diego. The usual caveat applies.

References

- Abeler, J., Becker, A. and Falk, A. (2014). Representative Evidence on Lying Costs, *Journal of Public Economics* 113: 96–104.
- Abeler, J., Nosenzo, D. and Raymond, C. (2019). Preferences for truth-telling, *Econometrica* 87(4): 1115–1153.
- Akers, R. (2017). *Social learning and social structure: A general theory of crime and deviance*, Routledge.
- Alesina, A. and La Ferrara, E. (2002). Who Trusts Others?, *Journal of Public Economics* 85(2): 207–234.
- Azar, O. H., Yosef, S. and Bar-Eli, M. (2013). Do Customers Return Excessive Change in a Restaurant?: A Field Experiment on Dishonesty, *Journal of Economic Behavior & Organization* 93: 219–226.
- Barber, B. M. and Odean, T. (2001). Boys Will Be Boys: Gender, Overconfidence, and Common Stock Investment, *The Quarterly Journal of Economics* 116(1): 261–292.
- Belot, M. and Van De Ven, J. (2017). How Private Is Private Information? The Ability to Spot Deception in an Economic Game, *Experimental Economics* 20(1): 19–43.
- Bersoff, D. (1999). Why Good People Sometimes Do Bad Things: Motivated Reasoning and Unethical Behavior, *Personality and Social Psychology Bulletin* 25: 28–39.
- Biziou-Van-Pol, L., Haenen, J., Novaro, A., Occhipinti Liberman, A. and Capraro, V. (2015). Does Telling White Lies Signal Pro-Social Preferences?, *Judgment and Decision Making* 10(6): 538–548.
- Broidy, L. and Agnew, R. (1997). Gender and Crime: A General Strain Theory Perspective, *Journal of Research in Crime and Delinquency* 34(3): 275–306.
- Buccioli, A., Landini, F. and Piovesan, M. (2013). Unethical Behavior in the Field: Demographic Characteristics and Beliefs of the Cheater, *Journal of Economic Behavior & Organization* 93: 248–257.

- Buccioli, A. and Piovesan, M. (2011). Luck or Cheating? A Field Experiment on Honesty With Children, *Journal of Economic Psychology* 32(1): 73–78.
- Burton Jr, V. S., Cullen, F. T., Evans, T. D., Alarid, L. F. and Dunaway, R. G. (1998). Gender, Self-Control, and Crime, *Journal of Research in Crime and Delinquency* 35(2): 123–147.
- Cappelen, A. W., Sørensen, E. Ø. and Tungodden, B. (2013). When Do We Lie?, *Journal of Economic Behavior & Organization* 93: 258–265.
- Capraro, V. (2018). Gender Differences in Lying in Sender-Receiver Games: A Meta-Analysis, *Judgment and Decision Making* 13(4): 345–355.
- Charness, G. and Dufwenberg, M. (2006). Promises and Partnership, *Econometrica* 74(6): 1579–1601.
- Charness, G. and Gneezy, U. (2008). What’s in a Name? Anonymity and Social Distance in Dictator and Ultimatum Games, *Journal of Economic Behavior & Organization* 68(1): 29–35.
- Charness, G., Haruvy, E. and Sonsino, D. (2007). Social Distance and Reciprocity: An Internet Experiment, *Journal of Economic Behavior & Organization* 63(1): 88–103.
- Chen, J. and Houser, D. (2017). Promises and Lies: Can Observers Detect Deception in Written Messages, *Experimental Economics* 20(2): 396–419.
- Childs, J. (2012). Gender Differences in Lying, *Economics Letters* 114(2): 147–149.
- Cohn, A., Gesche, T. and Maréchal, M. A. (2018). Honesty in the Digital Age, *CESifo Working Paper* 6996.
- Conrads, J. and Lotz, S. (2015). The Effect of Communication Channels on Dishonest Behavior, *Journal of Behavioral and Experimental Economics* 58: 88–93.
- Coricelli, G., Rusconi, E. and Villeval, M. C. (2014). Tax Evasion and Emotions: An Empirical Test of Re-Integrative Shaming Theory, *Journal of Economic Psychology* 40: 49–61.

- Croson, R. and Gneezy, U. (2009). Gender Differences in Preferences, *Journal of Economic Literature* 47(2): 448–474.
- Dai, Z., Galeotti, F. and Villeval, M. C. (2018). Cheating in the Lab Predicts Fraud in the Field: An Experiment in Public Transportation, *Management Science* 64(3): 1081–1100.
- Di Cagno, D., Galliera, A., Güth, W., Pace, N. and Panaccione, L. (2016). Make-Up and Suspicion in Bargaining With Cheap Talk: An Experiment Controlling for Gender and Gender Constellation, *Theory and Decision* 80(3): 463–471.
- Dreber, A. and Johannesson, M. (2008). Gender Differences in Deception, *Economics Letters* 99(1): 197–199.
- Drichoutis, A. C., Lusk, J. L. and Nayga, R. M. (2015). The Veil of Experimental Currency Units in Second Price Auctions, *Journal of the Economic Science Association* 1(2): 182–196.
- Dufwenberg, M. and Dufwenberg, M. A. (2018). Lies in Disguise—A Theoretical Analysis of Cheating, *Journal of Economic Theory* 175: 248–264.
- Dwenger, N. and Lohse, T. (2019). Do individuals successfully cover up their lies? evidence from a compliance experiment, *Journal of Economic Psychology* 71: 74–87.
- Erat, S. and Gneezy, U. (2012). White Lies, *Management Science* 58(4): 723–733.
- Ezquerro, L., Kolev, G. I. and Rodriguez-Lara, I. (2018). Gender Differences in Cheating: Loss vs. Gain Framing, *Economics Letters* 163: 46–49.
- Fosgaard, T. R., Hansen, L. G. and Piovesan, M. (2013). Separating Will From Grace: An Experiment on Conformity and Awareness in Cheating, *Journal of Economic Behavior & Organization* 93: 279–284.
- Gino, F., Krupka, E. L. and Weber, R. A. (2013). License to Cheat: Voluntary Regulation and Ethical Behavior, *Management Science* 59(10): 2187–2203.
- Gneezy, U., Gravert, C., Saccardo, S. and Tausch, F. (2017). A Must Lie Situation—Avoiding Giving Negative Feedback, *Games and Economic Behavior* 102: 445–454.

- Gneezy, U., Kajackaite, A. and Sobel, J. (2018). Lying Aversion and the Size of the Lie, *American Economic Review* 108(2): 419–53.
- Greiner, B. (2015). Subject Pool Recruitment Procedures: Organizing Experiments With ORSEE, *Journal of the Economic Science Association* 1(1): 114–125.
- Grosch, K. and Rau, H. A. (2017). Gender Differences in Honesty: The Role of Social Value Orientation, *Journal of Economic Psychology* 62: 258–267.
- Higgins, G. E. (2004). Gender and Self-Control Theory: Are There Differences in the Measures and the Theory’s Causal Model?, *Criminal Justice Studies* 17(1): 33–55.
- Holm, H. J. and Kawagoe, T. (2010). Face-To-Face Lying - An Experimental Study in Sweden and Japan, *Journal of Economic Psychology* 31(3): 310–321.
- Holt, C. A. and Laury, S. K. (2002). Risk Aversion and Incentive Effects, *American Economic Review* 92(5): 1644–1655.
- Houser, D., List, J. A., Piovesan, M., Samek, A. and Winter, J. (2016). Dishonesty: From Parents to Children, *European Economic Review* 82: 242–254.
- Houser, D., Vetter, S. and Winter, J. (2012). Fairness and Cheating, *European Economic Review* 56(8): 1645–1655.
- Jacobsen, C., Fosgaard, T. R. and Pascual-Ezama, D. (2018). Why Do We Lie? A Practical Guide to the Dishonesty Literature, *Journal of Economic Surveys* 32(2): 357–387.
- Kajackaite, A. and Gneezy, U. (2017). Incentives and Cheating, *Games and Economic Behavior* 102: 433–444.
- Khalmetski, K. and Sliwka, D. (2019). Disguising Lies-Image Concerns and Partial Lying in Cheating Games, *American Economic Journal: Microeconomics* 11: 79–110.
- Kocher, M. G., Schudy, S. and Spantig, L. (2018). I lie? we lie! why? experimental evidence on a dishonesty shift in groups, *Management Science* 64(9): 3995–4008.

- Konrad, K. A., Lohse, T. and Qari, S. (2014). Deception Choice and Self-Selection - The Importance of Being Earnest, *Journal of Economic Behavior & Organization* 107A: 25–39.
- Konrad, K. A., Lohse, T. and Qari, S. (2017). Compliance With Endogenous Audit Probabilities, *The Scandinavian Journal of Economics* 119(3): 821–850.
- Lohse, T. and Qari, S. (2018). Video Recordings in Experiments – Are There Effects on Self-Selection or the Outcome of the Experiment?, *Economics Bulletin* 38(3): 1381–1394.
- Lundquist, T., Ellingsen, T., Gribbe, E. and Johannesson, M. (2009). The Aversion to Lying, *Journal of Economic Behavior & Organization* 70(1-2): 81–92.
- Mears, D. P., Ploeger, M. and Warr, M. (1998). Explaining the Gender Gap in Delinquency: Peer Influence and Moral Evaluations of Behavior, *Journal of Research in Crime and Delinquency* 35(3): 251–266.
- Muehlheusser, G., Roider, A. and Wallmeier, N. (2015). Gender Differences in Honesty: Groups Versus Individuals, *Economics Letters* 128: 25–29.
- Niederle, M. (2016). Gender. *Handbook of Experimental Economics*.
- Pate, J. (2018). Temptation and cheating behavior: Experimental evidence, *Journal of Economic Psychology* 67: 135–148.
- Rebellon, C. J., Wiesen-Martin, D., Piquero, N. L., Piquero, A. R. and Tibbetts, S. G. (2015). Gender Differences in Criminal Intent: Examining the Mediating Influence of Anticipated Shaming, *Deviant Behavior* 36(1): 17–41.
- Roth, A. E. (2015). Is Experimental Economics Living Up to Its Promise?, *Handbook of Experimental Economic Methodology* pp. 13–40.
- Ruffle, B. J. and Tobol, Y. (2014). Honest on Mondays: Honesty and the Temporal Separation Between Decisions and Payoffs, *European Economic Review* 65: 126–135.
- Serra-Garcia, M. and Gneezy, U. (2019). Mistakes and Overconfidence in Detecting Lies, *mimeo* .

- Tibbetts, S. G. (1999). Differences Between Women and Men Regarding Decisions to Commit Test Cheating, *Research in Higher Education* 40(3): 323–342.
- Vanberg, C. (2008). Why Do People Keep Their Promises? An Experimental Test of Two Explanations, *Econometrica* 76(6): 1467–1480.
- Ward, S. J. and King, L. A. (2018). Gender Differences in Emotion Explain Women’s Lower Immoral Intentions and Harsher Moral Condemnation, *Personality and Social Psychology Bulletin* 44(5): 653–669.

Supporting information

The following regression uses women in the SI treatment as the unobserved reference category. It shows that women change their behavior significantly between treatment BL and SI, but not between SI and SIH.

Table S1: Probability of reporting honestly (SI female is reference)

	(1)
BL	-1.572** (0.748)
SIH	-0.842 (0.774)
Male	-1.750** (0.843)
Male * BL	0.985 (1.145)
Male * SIH	4.101*** (1.248)
Constant	1.583*** (0.575)
σ_u^2	14.47*** (3.725)
N	1155
Log pseudo lik.	-564.4
Chi-squared	20.44
prob>Chi2	0.00103

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

This table presents the results from a random effects logit model that predicts the individual probability of honestly reporting a high income. The predictors are dummies indicating the treatments (BL and SIH; SI is the omitted reference category), a dummy indicating male subjects and interaction terms. Random effects on the subject-level are included to take into account the repeated observations for the same subjects.

Summary: The results are qualitatively identical to the main results using the entire dataset. The results show that women change their behavior significantly between treatment BL and SI (the omitted reference category), but not between SI and SIH. This suggests that women react to the mode of communication (face-to-face vs. anonymous communication, but do not react to the strategic effects of perceived honesty.

The following regression uses only the data from the first round and employs women in the SI treatment as the unobserved reference category. It shows that women change their behavior significantly between treatment BL and SI, but not between SI and SIH.

Table S2: First round behavior: Probability of reporting honestly – logit model

	(1)
SI	-0.0547 (0.564)
SIH	1.347* (0.692)
Female	0.0253 (0.525)
SI * Female	1.020 (0.760)
SIH * Female	-1.096 (0.831)
Constant	0.262 (0.422)
<i>N</i>	192
pseudo R^2	0.040
Log pseudo lik.	-119.3
Chi-squared	8.760
prob>Chi2	0.119

Robust standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

This table presents the results from a logit model that predicts the individual probability to report honestly a high income using only the first round data. The predictors are dummies indicating the treatments (SI and SIH; BL is the omitted reference category), a dummy indicating female subjects and interaction terms.

Summary: The results are qualitatively identical to the main results using the entire dataset. The results show that men do not change their reporting behavior between BL (the omitted baseline category) and SI, i.e., men do not react to the psychological effects of face-to-face communication. Men’s truthfulness increases strongly in treatment SIH, i.e., men react strongly to the strategic effect of perceived honesty. Women’s truthfulness is similar in treatments SI and SIH. Furthermore, women’s truthfulness in SI and SIH is larger than in BL. This suggests that women react to the mode of communication (face-to-face vs. anonymous communication, but do not react to the strategic effects of perceived honesty).

Fig S1: Descriptives first round behavior – fraction of honest reports by gender

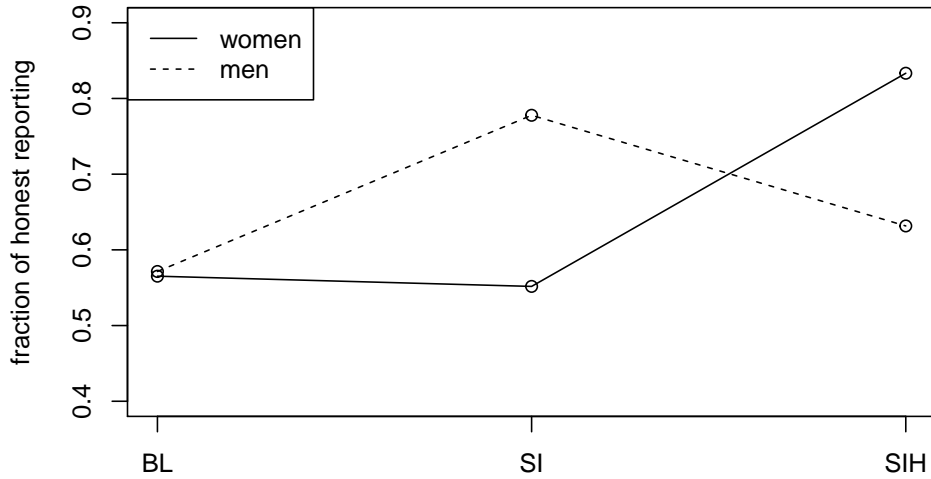


Table S3: First round behavior: Probability of reporting honestly (SI female is reference)

	(1)
BL	-0.965* (0.509)
SIH	-0.714 (0.525)
Male	-1.045* (0.549)
Male * BL	1.020 (0.760)
Male * SIH	2.116** (0.847)
Constant	1.253*** (0.402)
<i>N</i>	192
pseudo R^2	0.040
Log pseudo lik.	-119.3
Chi-squared	8.760
prob>Chi2	0.119

Robust standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

This table presents the results from a logit model that predicts the individual probability of honestly reporting a high income using the first round data. The predictors are dummies indicating the treatments (BL and SIH; SI is the omitted reference category), a dummy indicating male subjects and interaction terms.

Summary: The results show that women change their behavior significantly between treatment BL and SI (the omitted reference category), but not between SI and SIH. This suggests that women react to the mode of communication (face-to-face vs. anonymous communication), but do not react to the strategic effects of perceived honesty.

Table S4: Experiment II: Probability that a video is assessed as honest – logit model

	(1)
Female video	-0.0397 (0.122)
Female assessor	-0.0672 (0.0545)
Female video * female assessor	0.190** (0.0792)
Constant	0.0259 (0.0948)
<i>N</i>	9280
Log pseudo lik.	-6428.2
Chi-squared	6.146
prob>Chi2	0.105

Notes: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$, robust standard errors clustered on the level of videotaped subjects in parentheses. This table presents the results from a logit model that predicts that a video is assessed as honest.

Summary: The results show that the average honesty score (the probability that a video is assessed as honest) of videotaped women is larger if women are assessed by women, i.e., women are assessed as more honest by women. Men are assessed similarly by women and men.