



Working Papers

www.cesifo.org/wp

Do Individuals Put Effort into Lying? Evidence from a Compliance Experiment

Nadja Dwenger
Tim Lohse

CESIFO WORKING PAPER NO. 5805
CATEGORY 13: BEHAVIOURAL ECONOMICS
MARCH 2016

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

ISSN 2364-1428

Do Individuals Put Effort into Lying? Evidence from a Compliance Experiment

Abstract

We study whether individuals in a face-to-face situation can successfully exert some lying effort to delude others. We exploit data from a laboratory experiment in which participants were asked to assess videotaped statements as being rather truthful or untruthful. The statements are face-to-face tax declarations. The video clips feature each subject twice making the same declaration. But one time the subject is reporting truthfully, the other time willingly untruthfully. This allows us to investigate within-subject differences in trustworthiness. We find that a subject is perceived as more trustworthy if she deceives than if she reports truthfully. It is particularly individuals with dishonest appearance who manage to increase their perceived trustworthiness by up to 15 percent. This is evidence of individuals successfully exerting lying effort.

JEL-Codes: C910, H310, K420, H260.

Keywords: lying effort, face-to-face communication, perceived trustworthiness, lie detection, compliance.

Nadja Dwenger
University of Hohenheim
Schloss Hohenheim
Germany – 70593 Stuttgart
Nadja.Dwenger@uni-hohenheim.de

*Tim Lohse**
Berlin School of Economics and Law
Badensche Straße 52
Germany – 10825 Berlin
Tim.Lohse@hwr-berlin.de

*corresponding author

March 2, 2016

We thank Kai A. Konrad for valuable suggestions and a careful discussion of the experimental results. We also thank Uri Gneezy, Marta Serra-Garcia, and Julie Berry Cullen as well as seminar audiences at the Berlin School of Economics and Law and the WHU - Otto Beisheim School of Management for helpful comments and discussions. For providing laboratory resources we kindly thank MELESSA of the University of Munich and TU-Lab of the Technical University of Berlin. We thank Hans Mueller for developing and programming the web-based environment. Tim Lohse is grateful for hospitality at the University of California, San Diego, Rady School of Management, and for financial support of the Fritz Thyssen Foundation. The usual caveat applies.

1 Introduction

To deceive or not to deceive is a question that arises in basically all spheres of life. Sometimes the stakes involved are small and coming up with a lie is hardly worth it. But sometimes putting effort into lying might be rewarding – provided the deception is not detected. Whether or not a lie is detected is a matter of how trustworthy the individual is perceived to be. When interacting face-to-face two aspects determine the perceived trustworthiness:¹ first, an individual’s general appearance, and second, the level of some kind of effort the individual may choose when trying to make the lie appear truthful. Our experimental design offers an opportunity to make progress on the topic of deception and deception detection. We provide novel insights into the determinants of assessed trustworthiness and whether individuals actually can successfully exert lying effort. More specifically, we ask: do we really perceive individuals who tell the truth as more trustworthy than individuals who deceive? This is a fundamental question with a wide range of applications in real life lying situations. Think of a used car dealer ensuring that the deal of the day is not a lemon (although it is), the student claiming that he wrote the excellent term paper all by himself (although he didn’t), or the husband telling his wife that he is working late again (although he won’t). Despite its importance for social life, the literature has remained surprisingly silent on the issue of lying effort. This paper is the first to shed light on this issue.

We employ data from two types of experiments: an experiment with room for deception which was framed as a tax compliance experiment and a deception-assessment experiment. In the compliance experiment subjects had to declare income in face-to-face situations vis-à-vis an officer, comparable to the situation at customs. They could report honestly or try to evade taxes by deceiving. Some subjects received an audit and the audit probabilities were influenced by the tax officer, based on his impression of the subject. The compliance interviews were videotaped and some of these video clips were the basis for our deception-assessment experiment: For each subject we selected two videos both showing the same low income declaration, but once when telling the truth and once when lying. A different set of participants was asked to watch the video clips and assess whether the recorded subject was truthfully reporting her income or whether she was lying. These assessments were incentivised. Based on more than 18,000 assessments we are able to generate a trustworthiness score for each video clip (number of times the video is rated "rather truthful" divided by the total number of assessments). As each individual is assessed in two different video clips, we can exploit *within*-subject differences in trustworthiness. This allows us to condition our estimations on the general appearance (and the unobserved mindset) of an individual. Any difference in trustworthiness scores between situations of honesty and dishonesty can thus be traced back to the effort exerted by an

¹Face-to-face communication is the predominant form of communication for serious matters. Drawing on evidence from a survey which the Harvard Business Review conducted among its subscribers (Harvard Business Review Analytic Services 2009), Belot and van de Ven (2014) point out the importance of face-to-face meetings as crucial for building long-term business relationships (reported by 95% of the subscribers), for negotiating contracts (89%) or for understanding important clients (69%).

individual when lying. In addition, we also investigate whether subjects appear less trustworthy if they were audited and had been caught lying shortly before. What is important to note is that the individuals who had to assess the trustworthiness of a tax declarer did not receive any information on previous audits. That is, they could only base their assessment on the appearance and behaviour of the tax declarer in the videotaped situation that they were watching. We can thus study in isolation how catching individuals lying affects their trustworthiness through their own subsequent behaviour (e.g., because of a change in self-confidence) – beyond any signals that the successful exposure of lies additionally entail in other settings.

Our central results are as follows: Subjects appear as more trustworthy in compliance interviews in which they underreport than in compliance interviews in which they report truthfully. When categorizing individuals in subjects with a genuine dishonest or honest appearance, it becomes obvious that it is mainly individuals of the former category who appear more trustworthy when deceiving. These individuals with a dishonest appearance are able to increase their perceived trustworthiness by up to 15 percent. This finding is in line with the hypothesis that players with a comparably dishonest appearance, when lying, expend effort to appear truthful. Finding evidence for lying effort is new to the literature on deception and deception detection. We also find that an individual's trustworthiness is affected by previous audit experiences. Individuals who were caught cheating in the previous period, appear significantly less trustworthy, compared to individuals who were either not audited or who reported truthfully. This effect is exacerbated for individuals with a dishonest appearance if the individual is again underreporting but is lessened if the individual is reporting truthfully. No other paper has investigated into these effects before.

We proceed as follows. Section 2 briefly reviews the related literature. Section 3 explains the experimental set-up and provides a conceptual framework. Section 4 describes the data. In section 5 we present our empirical results and discuss them, and section 6 concludes.

2 Related Literature

Our results contribute to the literature on deception. Recent papers have experimentally studied a variety of aspects such as the importance of consequences from lying (Gneezy 2005), guilt aversion (Charness and Duwenberg 2006), the phenomena of overcommunication due to tensions between normative social behaviour and incentives for lying (Cai and Wang 2006, Sánchez-Pagés and Vorsatz 2007), an aversion to lying (Lundquist, Ellingsen, Gribbe and Johannesson 2009), behavioural differences depending on the type of lie (Erat and Gneezy 2012), the positive relationship between creativity and dishonesty (Gino and Ariely 2012), or the fact that people occupy a spectrum of preferences for truthfulness rather than only two opposite positions (Gibson, Tanner and Wagner 2013).² Despite the huge range of aspects analysed in the literature,

²One of the few theoretical papers is by Crawford (2003). He examines the advantage of a strategic lie by focusing on the interaction between rational and boundedly rational types of players with one-sided preplay costless communication.

all of these experimental studies about deception share a common feature of standard laboratory experiments: participants remain anonymous to each other and interaction and communication happen only via a computer. A crucial novelty compared to the previous literature is that both of these aspects are different in our paper: in the compliance experiment we allow for face-to-face interaction and direct communication with the officer. Our setup allows us to compare subjects in two face-to-face interaction situations, once when deceiving and once when being honest. We consider this an important difference to the existing literature as strategies of deceptive behaviour may not be invariant to the setting of communication. More specifically, the question of whether individuals exert some lying effort is more meaningful in situations with face-to-face communication with a person (our paper) than in a setting without face-to-face contact (standard in the deception literature).

We also contribute to the literature on deception detection which has for a long time exclusively been the sphere of social psychologists. This literature shows that a human's ability to identify liars is limited (for surveys, see DePaulo, Lindsay, Malone, Muhlenbruck, Charlton and Cooper 2003, Bond and DePaulo 2006, Vrij 2008, Hartwig and Bond 2011). Yet, DePaulo, Lindsay, Malone, Muhlenbruck, Charlton and Cooper (2003, p. 106) pointed at three shortcomings in much of the literature: study participants are not incentivised,³ participants are typically forced to lie instead of choosing for themselves whether to lie or not, and participants do not receive feedback on their deception performance. These weaknesses in the experimental design make it impossible to know whether results are driven by the carelessness of the participants (Smith 1994) and hard to investigate if participants exert some lying effort. Our experimental set-up allows us to overcome all of these issues.

Only recently the problem of detecting deception and assessing a subject's trustworthiness has found itself in the spotlight of economic research. An experiment which studies signals of dishonest communication is by Chen and Houser (2013). While their paper focuses on cues of dishonesty in written communication (such as the number of words used and the mention of money), our paper studies face-to-face communication. Such experimental work on whether individuals can unveil incomplete information in a situation that involves face-to-face communication has only been carried out by a few papers. Holm's and Kawagoe's (2010) experiment resembles a matching-pennies game in which players earn money if they can correctly assess whether their counterpart is lying or telling the truth. In their set-up, this counterpart has an incentive to choose a mixed strategy and theory predicts an equal mix of lying and truth-telling in equilibrium. In the actual experiment, however, they find that subjects have a bias towards truth telling when interacting face-to-face. Also, Konrad, Lohse and Qari (2016) observe an increase in truth-telling when income reporting does not happen anonymously through a computer but face-to-face towards a person with the power to decide whom to

³A recent partial exception is the work by Van Swol, Malhotra and Braun (2012) about detection of deception between either friends or strangers. However, while it is monetarily beneficial to deceive in their experiment, there is no sanction imposed on deceptive behavior if uncovered as it is in our study.

audit. The focus of the present paper is different from these studies in that we are focusing on how individuals are *perceived* when (not) telling the truth, while previous work have examined an individual's decision to deceive. Belot, Bhaskar and van de Ven (2012) experimentally investigate third-party predictions of subjects' behaviour in a prisoner's dilemma played on a TV show. Participants of the experiment saw (parts of) the show lasting from 6 up to 25 minutes. Drawing on rich communication between the players in the show, experimental participants could correctly predict players' trustworthiness. Our setting differs from Belot, Bhaskar and van de Ven (2012) in that the video taped communication is comparably short (only about 20 seconds) and followed a strict communication protocol. As shown by Serra-Garcia, van Damme and Potters (2011) lying behavior depends on whether subjects must send a precise message or are allowed to be vague. Using a standardized communication, we by construction exclude the various cues that Belot, Bhaskar and van de Ven (2012) identified, as these cues would be confounders of the lying effort which is the focus of our paper. Belot and van de Ven (2014) provide experimental evidence on the ability to detect deceit in a buyer-seller game with asymmetric information and face-to-face encounters. In contrast to our setting, Belot and van de Ven (2014) again allow free format communication which brings along many dimensions for deception detection. Our paper shows that in a comparably short and standardised communication sequence chances for detection are considerably lower. In fact, because individuals exert lying effort deceiving individuals even appear more trustworthy than honest individuals do. This leads to a detection rate below chance. The paper closest to ours is by Konrad, Lohse and Qari (2014). In the specific context of tax compliance with face-to-face declarations, they study deception choices and deception detection when individuals choose to lie according to whether they are perceived as honest or dishonest. Subjects who are perceived as honest are found to be more likely to deceive, leading to overall detection rates below chance. Whereas Konrad, Lohse and Qari (2014) observed each person just once and, therefore, applied a between-subject design, we observe each person twice, namely when behaving honestly and when behaving dishonestly. It is this *within*-subject design which crucially distinguishes our paper from Konrad, Lohse and Qari (2014) but also from Belot, Bhaskar and van de Ven (2012) and Belot and van de Ven (2014). The *within*-subject design allows us to answer our research question on whether or not deceiving individuals (successfully) exert lying effort. By comparing the perceived trustworthiness of the very same individual in two types of situations (truthful and non-truthful behavior), we are able to disentangle appearance (trustworthiness related to an individual's innate face characteristics etc.) and performance (related to an individual's costs of lying and lying effort).

Finally, this paper also contributes to the vast literature on tax compliance (recently surveyed, e.g., by Slemrod and Weber 2012). Despite the large amount of work on compliance, there is very little evidence on the dynamics of tax compliance and on a taxpayer's response to past audits and fines – in particular, when audit probabilities vary with an individual's observed behaviour. With our experimental setting, we provide novel insights on these aspects. Audit probabilities in previous studies were fixed across individuals and were independent of behaviour; in these studies deceptive behaviour was found to increase post-audit

(‘bomb crater effect’,⁴ cf., e.g., Guala and Mittone 2005, Maciejovsky, Kirchler and Schwarzenberger 2007, Kastlunger, Kirchler, Mittone and Pitters 2009).⁵ Unlike in these studies, individual audit probabilities in our setting vary with behaviour as they are a function of tax officers’ judgments (only average audit probabilities are given). When individuals consider future face-to-face tax compliance decisions (vis-à-vis another tax officer who has not been informed about prior audits), they thus have to take into account how trustworthy they are going to be perceived after an audit. If individuals suspect themselves of showing a lack of self-confidence or authenticity shortly after being caught cheating, they might well decide not to run the risk of lying and turn more compliant. The nature of the decision problem in face-to-face compliance situations is thus fundamentally different from one where uniform auditing probabilities apply.

3 Experimental set-up

Our experimental set-up consists of two parts: a videotaped tax compliance experiment and a deception-assessment experiment using some of the video clips from the compliance experiment. The videotaped tax compliance experiment was originally carried out by Konrad, Lohse, and Qari (2014). They generated a video database of a total of 472 incentivised compliance interviews as part of a larger research project but only used a small subset of 80 videos for their study. For our deception-assessment experiment we draw on the full database in order to select 96 videos featuring 48 individuals for whom a truthful and as well as a non-truthful tax declaration was recorded. The compliance experiment and the deception-assessment experiments were run on two distinct subject pools. In the following, we will briefly describe each of the two parts in more detail.

3.1 The compliance experiment

The compliance experiment measures the compliance choices of taxpayers in a set-up with face-to-face communication with a tax officer. Taxpayers had the possibility to either report high or low income to the tax officer. On average, half of the individuals who reported low income were subject to an audit which was certain to uncover the true income. The tax officers decided on whom to audit so that the effective individual audit probability depended on the appearance and performance of the taxpayer in the interview. The compliance experiment was videotaped and recorded.

Timing of events. The participants are randomly assigned either low (400 experimental units, with

⁴The term derives from the behavior of soldiers during the First World War: they sheltered in bomb craters as they believed that bombs would not fall in the same place twice.

⁵Note that the empirical evidence from the field is inconclusive (see Erard 1992, Bergman and Nevarez 2006, Kleven Knudsen, Kreiner, Pedersen and Saez 2011). This might be due to the fact that some tax payers expect subjective audit probabilities to increase with an audit while others assume audit probabilities to be independent of prior auditing activities.

probability 0.2) or high (1,000 experimental units, with probability 0.8) taxable income. Each participant is seated in front of a computer screen on which the amount of taxable income is privately displayed. Then, the taxpayers are called up individually by a message on their screen and sent into a compliance room. The sequence in which the taxpayers are called up is determined randomly. There were two tax compliance rooms, allowing for two people at a time to be called up for a private interview. In the compliance room a tax officer (a research student assistant, paid a flatrate) greets the person, identifies the person by her identification number (preserving pseudonymity throughout), and then asks for an oral tax declaration. The behaviour of tax officers follows a strict protocol. Depending on the income declared, taxpayers either do not have to pay tax (low income) or have to pay a tax of 200 experimental units (high income).

Once all participants have declared their income, half of the low income declarations are subject to an audit. Participants know that tax officers influenced the choice of whom to audit, i.e., that individual audit probabilities might be higher or lower than the average audit probability, depending on appearance and performance.⁶ Audited taxpayers with high income have to pay the tax of 200, topped up by a fine.⁷ It is a dominant strategy for taxpayers with low income to always declare low income. Taxpayers with high income have to decide whether to pay the tax and receive 800 experimental units or venture a gamble which yields 1,000 experimental units if underreporting remains undetected and $1,000 - (\text{tax} + \text{fine})$ experimental units if they are caught cheating. After audits are executed, each taxpayer privately learns on her computer screen whether she was audited or not and how much income she earned net of taxes and fines.

Implementation and payoffs. Each subject participated in four independent rounds. The taxpayer received a new draw of own income and was confronted with a different tax officer in each of these rounds, such that each tax officer met with a taxpayer only once.⁸ After the end of the fourth round, subjects were asked to answer a post-experiment questionnaire (non-incentivised).

At the end of the experiment, one round was chosen at random to determine the payoff of the participants. The average payment was EUR 17.90 per participant (with a standard deviation of EUR 3.57), including a fixed show-up fee of EUR 4.

The experiment was conducted at the experimental laboratory of the University of Munich (MELESSA) in March 2012 with students who were recruited using the software ORSEE (Greiner 2003).⁹ Before the

⁶Tax officers had to assess a series of declarations. They were asked to rank taxpayers according to their perceived dishonesty. This ranking determined who among the underreporting tax payers would be audited. Tax officers did not get any feedback on whether their ranking was justified or not. Hence, they could not learn. For more details, see Konrad, Lohse and Qari (2016).

⁷Konrad, Lohse and Qari (2014) varied the fine. In half of all sessions, it was low (100 experimental units), in the other it was high (300).

⁸This prevents later rounds from being affected by the impressions of a tax officer in a previous round. Re-matching also rules out second-order beliefs on the experience effects of tax officers on the side of the tax payers. Of course, there might still be learning effects for tax payers, such that we control for round effects in our analysis.

⁹The face-to-face compliance setting of Konrad, Lohse and Qari (2014) resembles the third treatment of a tax compliance experiment conducted by Konrad, Lohse, and Qari (2016). The focus of the latter study was on the role of face-to-face

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. After reading the instructions, all participants had to complete a quiz in order to make sure that they fully understood the experimental set-up.

Generating the video database. All face-to-face tax declarations were videotaped and voice-recorded. A clip takes about 20 seconds. Participants were informed about the recording before signing up and had to give their written consent.¹⁰ Since each participant played four rounds, there are four video clips per tax declarer. In total, Konrad, Lohse and Qari (2014) generated a video database with 472 video clips.

Our selection of videos. Depending on a participant’s income and her declaration choice if she earned high income, a participant is potentially observed in three different situations: truthfully declaring high income, truthfully declaring low income, and underreporting high income. Useful for our deception-assessment experiment are the two latter types of video clips, in which individuals truthfully or untruthfully declare income to be low. In what follows, we focus on the subset of participants for whom at least one video clip shows a situation with a truthful report of low income and a situation with an untruthful report of low income (while true income was high). We select exactly one video clip per situation and person. This course of action allows us to make sure that later assessments of the video clips are not driven by general differences between truthfully and untruthfully reporting individuals.¹¹ The procedure leaves us with 48 subjects and 96 video clips from the video database.

3.2 The deception-assessment experiment

In the second part of the experiment, we invited a large number of participants to assess the truthfulness of some of the videotaped tax declarations. The assessments as ‘rather untruthful’ and ‘rather truthful’ reports allow us to generate a general trustworthiness score for each video clip as we describe below.

Samples of video clips shown and observations. We randomly divide the set of 96 clips into two disjoint subsets of 48 clips. Each subset contains 24 truthful and 24 untruthful declarations of low income and shows a taxpayer exactly once. Both subsets of video clips were shown to 190 assessors each, i.e., 380 subjects took part in the experiment (grouped in 16 sessions with up to 24 participants). Each subject assessed 48 video clips, which leaves us with 380 subjects times 48 assessments/subject = 18,240 assessments.

interaction for compliance choices, compared to an anonymous fully computerized tax compliance treatment.

¹⁰Participants were also truthfully told that the videotapes would be used for scientific purposes only, and that strict pseudonymity is preserved. Of course, participants were free to revoke their consent and to quit the experiment at any time being only paid the show-up fee. But no one has done that. Privacy is also kept in the deception-assessment experiment and there is no feedback provided to assessors on whether a specific clip shows a truthful low-income report or a high-income tax payer who underreports.

¹¹Always-liars and always-truth-tellers are not the focus of our study. Psychologists stress that healthy human beings regularly lie for various reasons but not in all interactions.

Timing of events. Each assessor sits in front of a computer screen on which the 48 video clips (one of the two subsets of clips) are shown in sequential but random order. Assessors have a headphone to listen to the taxpayer’s income declaration. After the display of the video clip the assessor is asked to make a binary assessment, i.e., to assess whether the person shown in the video clip is more likely to truthfully or to untruthfully report a low income. Assessors have to pass judgement on the clip displayed before moving on to the next video clip. There is no possibility to return and change the assessment.

Implementation and payoffs. The deception-assessment experiment took place at the experimental laboratory of the Technical University Berlin in November 2013. By recruiting participants for the compliance experiment in Munich and those for the deception-assessment experiment in Berlin, we tried to make sure that (i) there was no overlap of subjects and (ii) that the assessor did not personally know the videotaped individuals. Participants in Berlin were also recruited using the software ORSEE (Greiner 2003).

At the beginning of the deception-assessment experiment, detailed instructions (see Appendix) were read aloud and handed out in print. The instructions explained the task of assessing video clips, the sequential but random order of video clips, and the fact that each clip could only be assessed once. They also explicated that the ratio of truthful to untruthful reports in the set of video clips shown was about 50/50.¹²

At the end of the experiment, 5 out of 48 assessments were randomly selected to determine the payoff of the participants. Each video clip correctly assessed was worth EUR 4.50, i.e., the assessment task could earn participants up to EUR 22.50 (plus fixed show-up fee of EUR 5). On average, participants in the deception-assessment experiment earned EUR 16.17 (with a standard deviation of EUR 5.01), including the show-up fee.

Trustworthiness score. For each of the 96 video clips v shown we generate a trustworthiness score TWS_v for each respective clip v as

$$TWS_v = \frac{\# \text{ of times } v \text{ is rated "rather truthful"}}{\text{total } \# \text{ of assessments}} * 100\% \quad (1)$$

Two points are of note. First, recall that the deception-assessment experiment took place many months after the actual compliance experiment. Hence, taxpayers in the compliance experiment did not know their trustworthiness score when taking their tax compliance decisions. However, as elaborated below we expect taxpayers to self-evaluate whether they generally appear more or less trustworthy; this self-evaluation may be correlated with the external assessment captured in the trustworthiness score. Second, we calculate the trustworthiness score for each single video clip. Consequently, two scores are available for each taxpayer: one for the video clip which shows her truthfully declaring low income and another one for the video clip showing her underreporting. This within-person variation in trustworthiness is one of the key advantages of this study compared to the previous literature as it allows us to disentangle appearance and performance as determinants of trustworthiness.

¹²Assessors responded to this information by rating about half of the videos (48.71%) as rather untruthful in the experiment.

3.3 Conceptual framework

Forming a hypothesis on the questions i. whether an individual is perceived as more trustworthy if she is truthful than if she deceives, ii. whether individuals exert lying effort and iii. how this effort may influence an individual’s perceived trustworthiness is far from straightforward. In specifying the determinants of an individual’s trustworthiness, three factors are at play which potentially work in opposite directions: an individual’s general appearance, some individual cost of deception, and an individual’s lying effort.

First, whether we perceive somebody as trustworthy or not depends on an individual’s appearance. We are more likely to trust someone with an ‘honest face’ (Rezlescu, Duchaine, Olivola and Chater 2012). These facial features – just like a person’s genuine attitude toward deception – cannot be easily changed. Without loss of generality, individuals’ appearances can be categorized in comparably honest looking individuals ($A_i = H$) and dishonest looking individuals ($A_i = D$). A comparably more honest general appearance contributes to a higher trustworthiness score: for an honest looking individual i the effect on his trustworthiness score is denoted by $a_i^H \in [a^H, a^{\bar{H}}]$, and for a dishonest looking individual $a_i^D \in [a^D, a^{\bar{D}}]$ with $a^D \geq 0$, and $a^{\bar{D}} < a^H$.

Second, some psychological effects come along with a deceptive strategy and may affect an individuals perceived trustworthiness. The literature has identified mental cost of lying (Vanberg 2008, Lundquist, Ellingsen, Gribbe and Johannesson 2009), shame (Coricelli, Joffily, Montmarquette and Villeval 2010) or guilt (Charness and Dufwenberg 2006), among others, as relevant dimensions for deception decisions.¹³ In particular, pursuing a deceptive strategy is costly for the individual as it adversely affects an individual’s perceived trustworthiness when he blushes and shows other signs of stress. We denote whether an individual lies or not with $L_i = 1$ and $L_i = 0$, respectively, and summarize the negative consequences of lying for an individual’s perceived trustworthiness with $c_i(L_i)$ with $c_i(0) = 0$, and $c_i(1) = c_i > 0$.

Third, whether an individual can be caught lying depends on the effort she exerts in order to present herself as trustworthy. In the context of our compliance experiment, a subject who is truthful has no incentive to make any suchlike effort: a potential audit of her compliance statement has literally no consequences and involves no costs for her. In contrast, a subject who deceives has an incentive to exert effort to present herself as trustworthy and to avoid being caught. For instance, the subject may actively try to avoid any visible signals of nervousness or stress which could be used as cues by the auditor. Hence, if deceiving subjects exert a lot of effort on presenting themselves as trustworthy, they might even be assessed as more trustworthy than honest subjects are. $e_i(L_i)$ captures the individual lying effort with $e_i(1) = e_i \geq 0$ and $e_i(0) = 0$.

From a theory perspective, for individual i her trustworthiness score results from the different combina-

¹³These undesirable side effects of lying can also explain why not all individuals use lies to maximize their monetary payoffs (Gneezy 2005, Sutter 2009, Erat and Gneezy 2012, Fischbacher and Föllmi-Heusi 2013, Gino, Krupka and Weber 2013). Recent economic models embed the loss in utility from misrepresenting private information (e.g., Kartik 2009; Kartik, Tercieux and Holden 2014).

tions of general appearance as dishonest or honest and deception decision, $TWS_i(A_i, L_i)$:

$$TWS_i(D, 0) = a_i^D \quad (2)$$

$$TWS_i(D, 1) = a_i^D - c_i + e_i \quad (3)$$

$$TWS_i(H, 0) = a_i^H \quad (4)$$

$$TWS_i(H, 1) = a_i^H - c_i + e_i \quad (5)$$

The trustworthiness scores as described in (2) to (5) allow forming some hypotheses. By construction, $TWS_i(D, 0) < TWS_i(H, 0)$. Recall that individuals in the compliance experiments were not forced to lie. When deciding about whether to deceive or not an individual weighs the negative consequences of lying on her perceived trustworthiness against the effect of the lying effort to cover the lie. As individuals may well be aware of how honestly they are perceived by others,¹⁴ we conjecture that an individual would only take a deceptive strategy if she expects to appear at least as trustworthy as when she would make a truthful strategy. In other words, we expect an individual to lie whenever her lie remains undetected (due to a non-decreased trustworthiness score) and she benefits somewhat from her lie. Defining $\Delta_i^D := TWS_i(D, 1) - TWS_i(D, 0)$ and $\Delta_i^H := TWS_i(H, 1) - TWS_i(H, 0)$, we can state the following hypothesis:

Hypothesis 1 (Existence of lying effort)

- a) Individuals with dishonest appearance are able to exert some lying effort, i.e. $\Delta_i^D > 0$.
- b) Individuals with honest appearance are able to exert some lying effort, i.e. $\Delta_i^H > 0$.

Both inequalities provide evidence for individuals successfully exerting some lying effort.

It is a particular interesting question whether Δ_i^D is smaller, larger or equal to Δ_i^H . On the one hand, one might argue that individuals of type H are somewhat gifted and, hence, could be able to appear more easily as even more honest when trying to cover a lie. But on the other hand, H type individuals are already perceived as particularly honest, so that they can afford to appear a little less trustworthy – and still remain perceived as more honest than dishonest type individuals. Compared to D type individuals, H type individuals might hence spend less effort on appearing trustworthy when lying. In contrast, by definition even the most honest appearing individual of type D has a lower trustworthiness score than the least honest appearing individual of type H , i.e. $a^{\bar{D}} < a^{\underline{H}}$. This implies that D type individuals *per se* have a much bigger incentive to exert some lying effort:

Hypothesis 2 (Size of lying effort by type)

¹⁴Konrad, Lohse and Qari (2014) square self-assessed trustworthiness with the trustworthiness perceived by others and find that there is an extremely high correlation between those two.

Individuals with a dishonest appearance exert a larger lying effort than individuals with an honest appearance, i.e. $\Delta_i^D > \Delta_i^H$.

To test these hypotheses empirically, we use the empirical trustworthiness score as described in eq. (1). We classify individuals into D and H appearance types based on the trustworthiness score they achieve when declaring truthfully ($L_i = 0$) as given by equations (2) and (4), respectively. Individuals with a trustworthiness score below the median are categorized as dishonest appearing types ($A_i = D$), those with scores strictly above the median are honest appearing types ($A_i = H$). Using the classification of individuals from situations in which the individual makes an untruthful statement ($L_i = 1$) we can derive the trustworthiness scores of (3) and (5), respectively.

Finally, we can conjecture how trustworthy individuals are perceived who had been caught lying shortly before. Given the design of the compliance experiment with an endogenous audit probability, we expect an individual's trustworthiness score to be negatively affected by a previous lie detection. The reason for this is that the individual's self-confidence may have suffered severely. Hence, in light of her previous auditing experience, the individual may show even more signs of stress now. In order to capture these dynamic effects on an individual's perceived trustworthiness denote with P_i whether an individual had been caught lying previously ($P_i = 1$) or not ($P_i = 0$). While there is no effect on an individual's trustworthiness score in the latter case, in the former case an individual's score is assumed to decrease by $p_i > 0$. Taking the dynamic effects on individuals trustworthiness into account, the resulting trustworthiness scores $TWS_i(A_i, L_i, P_i)$ read:

$$TWS_i(D, 0, 0) = a_i^D \quad (6)$$

$$TWS_i(D, 0, 1) = a_i^D - p_i \quad (7)$$

$$TWS_i(D, 1, 0) = a_i^D - c_i + e_i \quad (8)$$

$$TWS_i(D, 1, 1) = a_i^D - c_i - p_i + e_i \quad (9)$$

$$TWS_i(H, 0, 0) = a_i^H \quad (10)$$

$$TWS_i(H, 0, 1) = a_i^H - p_i \quad (11)$$

$$TWS_i(H, 1, 0) = a_i^H - c_i + e_i \quad (12)$$

$$TWS_i(H, 1, 1) = a_i^H - c_i - p_i + e_i \quad (13)$$

The modified trustworthiness scores as described in (6) to (13) allow forming the following hypothesis:

Hypothesis 3 (Dynamics of being caught)

Individuals of dishonest and honest type who had been caught lying previously appear less trustworthy both when they are actually honest, i.e. $TWS_i(D, 0, 1) < TWS_i(D, 0, 0)$ and $TWS_i(H, 0, 1) <$

$TWS_i(H, 0, 0)$, respectively, and when they are deceiving, i.e. $TWS_i(D, 1, 1) < TWS_i(D, 1, 0)$ and $TWS_i(H, 1, 1) < TWS_i(H, 1, 0)$.

4 Data

Table 1 shows descriptive statistics for both parts of our experiment. Panel A refers to the characteristics of the individuals whose tax declarations were videotaped (thus ‘declarers’) in the first part of the experiment. Panel B displays descriptives of all those participants who were asked to assess video clips (‘assessors’) in the second part of the experiment.

Panel A shows that individuals whose tax declarations were videotaped and selected for later assessment were, on average, 22.6 years old. About 56% of them were female and about 16.7% of them indicated to have experienced customs control in real life.¹⁵ As explained above only when an individual had been observed both in a situation with truthful and untruthful reporting of income were they chosen for further assessment, and we selected exactly one video clip per situation and person. This implies that the share of untruthful reports in the video clips assessed is equal to 50% by construction. This figure compares to a share of untruthful reports of 75.7% (calculated as # of untruthful reports / # of situations with high income earned) if we look at all tax declarations made by the selected declarers.

Panel B shows the average characteristics for the 380 individuals who were asked to assess the selected video clips in the second part of the experiment. On average, assessors were 24.3 years old and about 42.6% of them were female. They assessed 49.7% of all video clips correctly. Note that this share is statistically different from the average score of 50%, which pure random choice (flipping a coin) would have given them. Figure 1 classes this result with the previous literature on deception detection tasks as reviewed by Bond and DePaulo (2006). They report on assessment scores for a large number of lie-catching experiments in social psychology, with an average of 53.46% correct lie-truth classifications.

Our findings differ from those in the literature in two aspects. Firstly, the share of correct assessments in our study is based on a huge number of assessments ($N = 18,240$). Secondly, our tax compliance experiment features a more sophisticated design which addresses the three major shortcomings of many previous studies as discussed earlier: First, the participants of both the compliance and the lie-catching experiment are incentivised as they earn money if they are successful. Second, all subjects who are seen on the videotapes are performing an action which they have chosen, based on their true taxable incomes, their monetary incentives to potentially deceive, and their perceptions about their (un)trustworthy appearance. This is an important departure, as the ‘quality of liars’ is different compared to a setting in which all subjects who lie

¹⁵Around a third of the video clips (31.3%) were recorded in sessions with a high fine.

Table 1: Descriptives

A: Declarer	
Age	22.625 (3.266)
Male	.438 (.501)
Experienced Customs Control	.167 (.377)
Participated in Treatment with High Fine	.313 (.468)
Share of Untruthful Reports in Assessment Sample (By Construction)	.500 (.000)
Share of Untruthful Reports in Full Sample	.757 (.268)
Number of Declarers in Assessment Sample	48

B: Assessor	
Age	24.250 (4.341)
Male	.574 (.495)
Share of Correct Assessments	.497 (.073)
Average Payment [in EUR]	16.167 (5.009)
Number of Assessors	380

Notes: Panel A shows average characteristics for the 48 declarers in our assessment sample. Panel B displays the average characteristics of the 380 assessors. Each assessor makes a judgement on the behavior of all 48 declarers in the assessment sample, which leads to a total number of 18,240 judgments in our sample (380 assessors x 48 judgments/assessor).

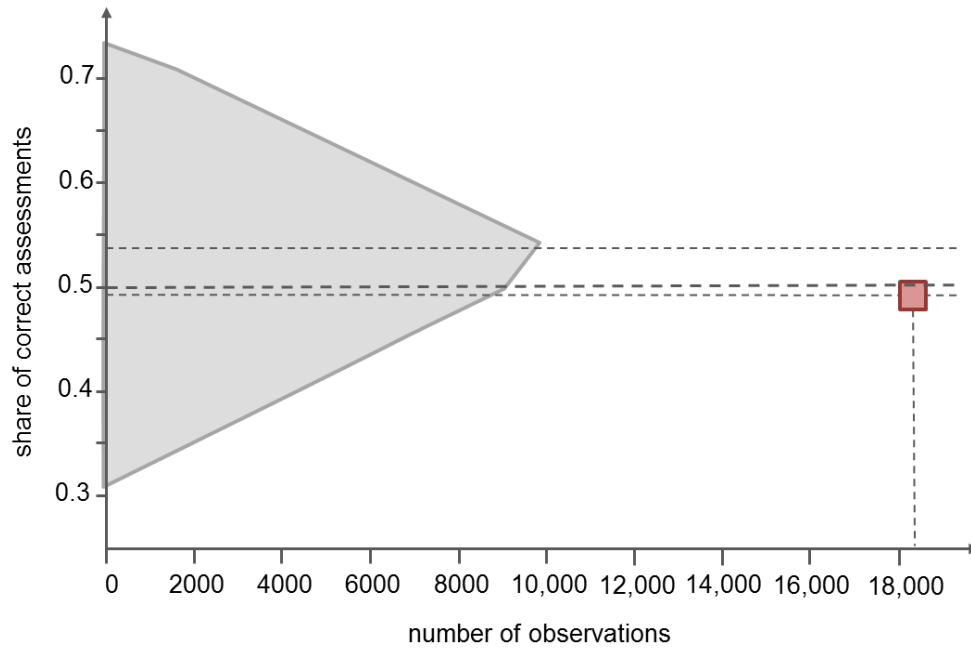


Figure 1: Spectrum of the average shares of correct assessments and number of observations in previous lie-catching studies compared to our results. The centre of the square locates the outcome of this paper. The grey shaded area is the convex hull of the findings reported by Bond and DePaulo (2006, p. 222) in their meta study. The upper dashed line shows the average share of correct assessments in these studies, the lower dashed line represents the share in this study (0.497). Source: Own depiction

are forced to, or are advised to lie. Studying lying effort requires that the analysis draws on evidence from a situation in which a subject freely decides to deceive compared to a situation in which the subject freely decides not to do so. Third, participants of the compliance experiment receive direct feedback since they face a risk of being detected every round they cheat. This also allows us to provide evidence on the dynamics of an individual’s perceived trustworthiness that unfold if the individual was caught lying in a previous round. We expect all of these differences in experimental design to increase the quality of untruthful reports and thus to deteriorate assessors’ lie-catching performance.

5 Results

This section contains our experimental results. First, we analyse whether taxpayers successfully exert effort when underreporting their income. Second, we look at the dynamics of being caught in order to understand whether successful audits have secondary effects, by making it more difficult for individuals to appear trustworthy right after being caught.

5.1 Trustworthiness and lying effort

We now analyse the hypothesis that individuals successfully exert additional effort in order to appear trustworthy when they are underreporting. To do so, we compare the trustworthiness scores assigned to video clips which show truthful declarations of low income to the trustworthiness scores assigned to video clips which show untruthful declarations of low income. As explained earlier, we calculate trustworthiness scores for all video clips as the number of assessments rating the behaviour ‘rather truthful’, divided by the total number of assessments. Table 2 presents the regressions results on average trustworthiness scores in both situations. Columns (1) to (3) show the full sample, while columns (4) to (9) split the sample into dishonest and honest appearance types. Standard errors are clustered at the level of the assessor (to take into account that each assessor evaluates several videos). Three points are of note.

First, the linear regressions for the full sample clearly show that trustworthiness scores are higher if taxpayers underreport. The baseline estimation in column (1) does not include any control variables, while the specification in column (2) includes assessor fixed effects and controls for observable characteristics of tax declarers, such as age and sex. Part of the videos were recorded in a high fine regime (fine of 300 instead of 100 experimental units), for which we also control. The specification in column (3) additionally includes a control variable for the round of the experiment to account for potential changes in behaviour of tax declarers over time through learning or fatigue. In the baseline estimation (1), the trustworthiness score is 0.51 for truthful reports and 0.52 for untruthful reports. The same picture arises for specifications (2) and (3). This implies that, on average, individuals are able to fool assessors by exerting lying effort –

Table 2: Lying Effort

Dependent Variable: Trustworthiness Score	Full Sample			Dishonest Appearance			Honest Appearance		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Baseline: Truthful Report	.510*** (.001)	.502*** (.001)	.516*** (.003)	.399*** (.001)	.419*** (.002)	.456*** (.001)	.630*** (.001)	.590*** (.002)	.557*** (.005)
Untruthful Report	.006*** (.001)	.006*** (.001)	.009*** (.001)	.059*** (.001)	.060*** (.002)	.062*** (.001)	-.051*** (.001)	-.050*** (.001)	-.064*** (.001)
Round			-.006*** (.001)			-.014*** (.001)			.017*** (.001)
With Tax Declarer and Assessor Observable Characteristics	no	yes	yes	no	yes	yes	no	yes	yes
Number of Observations	18 240	18 240	18 240	9 500	9 500	9 500	8 740	8 740	8 740
Number of Independent Observations	380	380	380	380	380	380	380	380	380
R ² (overall)	.001	.070	.072	.084	.176	.201	.053	.092	.113

Notes: *** denotes significance at 1% level. Standard errors clustered at the level of the assessor are given in parentheses. The unit of observation throughout is the video. The trustworthiness score is calculated for each videoclip as the number of assessments rating the behavior "truthful", divided by the total number of assessments. "Untruthful report" is a dummy variable (equal to one if the individual is deceiving and zero otherwise), and "baseline: truthful report" refers to the constant. We split the sample into "dishonest appearance" types (trustworthiness score below median for truthful declaration) and "honest appearance" types (trustworthiness score strictly above median for truthful declaration). Tax declarer and assessor observable characteristics in columns (2), (3), (5), (6), (8) and (9) consist of the age variables and sex indicators for tax declarers and assessors, and in a "high fine" indicator as part of the tax declarers made their declaration in a high fine regime. Each tax declarer was asked to take four declaration decisions during the experiment; the control variable "round" ranges from 1 (round 1) to 4 (round 4) and controls for any changes related to time such as fatigue or learning.

Summary: The results show that in the full sample the declaring individuals appear to be more trustworthy if they untruthfully declare their income, compared to a situation in which they truthfully declare their income. This suggests that the declaring individuals successfully exert effort in order to appear trustworthy. Splitting the sample reveals that this finding is fully driven by dishonest looking types. Dishonest looking types appear more trustworthy when lying. Honest looking types appear less trustworthy when lying. However, even when exerting lying effort the average dishonest looking type of person does not fully make up for her appearance as she still looks less trustworthy than an honest looking type of person does.

trustworthiness scores even increase beyond the benchmark score achieved for truthful declarations (relative increase of 1 to 2%).¹⁶

Second, the above finding is fully driven by the individuals with dishonest appearance as it can be seen in (4) to (6). Starting from low levels when reporting honestly, these individuals on average increase their trustworthiness scores by about 0.060 percentage points, or 12 to 15%. We interpret this finding as evidence for significant lying effort as conjectured in Hypothesis 1a).

Third, individuals with honest appearance enjoy high levels of trustworthiness by definition. They seem to exert much less effort of lying than individuals with dishonest appearance do (associated change in the trustworthiness score -0.5 vs. +0.6). The lying effort exerted by the group of honestly looking individuals is not enough to compensate for the loss in perceived trustworthiness associated with lying. From (7) to (9) it becomes obvious that, on average, lying individuals with honest appearance lose 8% of their trustworthiness compared to initial levels. This finding contradicts Hypothesis 1b).

Taken together, these findings support Hypothesis 2: the size of the lying effort is indeed larger for individuals with a dishonest appearance than for those with an honest appearance.

¹⁶Note that this result cannot be driven by self-selection of good or bad liars as our sample features the honest and dishonest behavior of the very same set of individuals.

We summarise the above finding in:

Result 1: Individuals deceive assessors successfully and appear more trustworthy when underreporting their income. This suggests that they expend effort to appear trustworthy. This results in particular holds true for individuals with dishonest appearance.

5.2 Trustworthiness and dynamics of being caught

Next, we check whether taxpayers' ability to deceive and to appear trustworthy in underreporting situations is influenced by whether they have previously been caught cheating or not.

Table 3 provides regression results on this issue. The dependent variable in all estimations is the trustworthiness score. Table 3 shows that subjects appear significantly less trustworthy if they had been caught lying in the previous period. In the full sample (column (1)), we find the average trustworthiness score for these individuals to be about 17% (or 0.082 trustworthiness score points) lower, compared to those for individuals who either truthfully reported in the previous period or went uncaught. Please note that this effect arises even though neither tax officers in the tax compliance experiment nor assessors in the deception-assessment experiment were informed of an individual's prior audits or declarations. The detrimental effect of being caught on subsequent trustworthiness is even more pronounced for those individuals underreporting (-32% or -0.143 trustworthiness score points, column (3)). This suggests that it is difficult for individuals to overcome the experience of having been caught and fined for underreporting in a previous round: if individuals rightly anticipate that they appear less trustworthy after being caught they should respond by reporting honestly. In our data we indeed find a negative (but statistically insignificant) effect of being caught on the probability of underreporting (point estimate: -0.133, p -value: 0.201). This finding contradicts the 'bomb crater effect' described in laboratory experiments with audit probabilities which were fixed across individuals and were independent of individual behaviour (see, e.g., Guala and Mittone 2005).

To dig deeper into the aftereffects of being caught, we investigate whether the loss in trustworthiness relates to an individual's appearance as being rather honest or dishonest. We measure dishonest appearance by a binary variable which is equal to one if a taxpayer's average trustworthiness score for the videoclip showing a truthful declaration of low income is smaller than (or equal to) the median trustworthiness score of all truthfully declaring individuals, and zero otherwise. As expected, we find the trustworthiness scores of individuals with a dishonest appearance to be below average (by construction, see columns (2) and (4)). What is interesting, though, is the interaction effect between dishonest appearance and 'caught in previous period'. While the interaction term is positive for the full sample (column (2)), it is negative if we only consider untruthful declarations (column (4)). This suggests that having been caught previously affects a taxpayer's self-confidence: Individuals with a dishonest appearance may present themselves as even more

Table 3: Dynamics of Being Caught

Dependent Variable: Trustworthiness Score	Full Sample		Untruthful Declarations	
	(1)	(2)	(3)	(4)
Baseline	.490*** (.010)	.529*** (.007)	.446*** (.012)	.450*** (.011)
Caught in Previous Period	-.082*** (.003)	-.057*** (.004)	-.143*** (.004)	-.039*** (.008)
Dishonest Appearance		-.173*** (.002)		-.100*** (.002)
Caught in Previous Period x Dishonest Appearance		.019*** (.005)		-.081*** (.009)
Controlling for Tax Declarer and Assessor Observable Characteristics	yes	yes	yes	yes
Number of Observations	18 240	18 240	9 120	9 120
R ²	.067	.434	.161	.317

Notes: *** denotes significance at 1% level. The unit of observation throughout is the video. Columns (1) and (2) display the results for the full sample and columns (3) and (4) display the results for the subsample of untruthful declarations. The trustworthiness score is calculated for each videoclip as the number of assessments rating the behavior "truthful", divided by the total number of assessments. Dishonest appearance of each declarer is measured by a binary variable which is equal to one if the declarer's average trustworthiness score in situations of truthful declaration is weakly smaller than the median trustworthiness score of all individuals in situations with truthful declaration, and zero otherwise. "Caught in previous period" is a binary variable which is equal to one if the declaration considered was made one period after the declarer had been caught making an untruthful declaration. Tax declarer and assessor observable characteristics in columns (2) and (4) consist of the age variables and sex indicators for tax declarers and assessors.

Summary: The results show that declarers are assigned a lower trustworthiness score if the declaration considered was made one period after the declarer had been caught making an untruthful revelation. Comparing columns (1) and (2) to columns (3) and (4) shows that the effect is strongest in situations of untruthful declarations, in particular for declarers with a dishonest appearance.

insecure and uncomfortable when again underreporting (leading to even lower scores of trustworthiness) but as relaxed and relieved when they are reporting truthfully (leading to higher scores of trustworthiness).

All in all, we find support for Hypothesis 3 and summarise the above in:

Result 2: Trustworthiness is affected by previous audit experiences. Individuals who were caught cheating in the previous period appear significantly less trustworthy, compared to individuals who were either not audited or who reported truthfully. This effect is exacerbated for individuals with a dishonest appearance if the individual is again underreporting but is lessened if the individual is reporting truthfully.

6 Conclusion

The present paper experimentally studies whether individuals in a face-to-face situation can successfully exert some lying effort to delude others. We exploit data from a laboratory experiment in which participants were asked to assess videotaped statements as being rather truthful or untruthful. The statements are face-to-face tax declarations which were recorded in an incentivised tax compliance experiment. The video clips to be assessed feature each taxpayer twice with the same declaration. But one time the subject is reporting truthfully, the other time willingly untruthfully. This allows us to investigate within-subject differences in trustworthiness. Drawing on more than 18,000 assessments, we find that a subject is perceived as more trustworthy if she deceives than if she reports truthfully. We categorized individuals in subjects with a genuine dishonest or honest appearance. Whereas individuals of the latter category have only little incentive to cover their lies as they appear comparably honest anyways, it is mainly individuals of the former category who appear more trustworthy when deceiving. These individuals with a dishonest appearance are able to increase their perceived trustworthiness by up to 15 percent. This represents the first conclusive empirical evidence of individuals successfully exerting an effort to lie. Moreover, our results show that an individual appears less trustworthy shortly after she is caught lying – even though the person assessing the truthfulness of the tax declaration had not been informed about prior audits. Anticipating this loss in trustworthiness, individuals should immediately turn more compliant once they have been caught cheating. This makes the compliance decision problem in face-to-face situations (in which audit probabilities depend on perceived trustworthiness) very different from the one with fixed, behaviour-independent audit probabilities. Our data provide suggestive evidence that individuals indeed are less likely to cheat after being caught, which contrasts the well-described post-audit decrease in compliance (‘bomb crater effect’) for fixed audit probabilities.

References

- [1] Belot, Michèle, V. Bhaskar, and Jeroen van de Ven, 2012, Can observers predict trustworthiness, *Review of Economics and Statistics*, 94(1), 246–259.
- [2] Belot, Michèle, and Jeroen van de Ven, 2014, How private is private information? The ability to spot deception in an economic game.
- [3] Bergman, Marcelo and Armando Nevarez, 2006, Do audits enhance compliance? An empirical assessment of VAT enforcement, *National Tax Journal*, 59(4), 817–832.
- [4] Bond, Charles F. and Bella M. DePaulo, 2006, Accuracy of deception judgments, *Personality and Social Psychology Review*, 10(3), 214–234.
- [5] Cai, Hongbin, and Joseph Tao-Yi Wang, 2006, Overcommunication in strategic information transmission games, *Games and Economic Behavior*, 56, 7–36.
- [6] Charness, Gary, and Martin Dufwenberg, 2006, Promises and partnership, *Econometrica*, 74, 1579–1601.
- [7] Chen, Jingnan Cecilia, and Daniel Houser, 2013, Promises and Lies: An Experiment on Detecting Deception. *Working Paper* No. 1038.
- [8] Coricelli, Giorgio, Mateus Joffily, Claude Montmarquette, and Marie Claire Villeval, 2010, Cheating, emotions, and rationality: an experiment on tax evasion, *Experimental Economics*, 13(2), 226–247.
- [9] Crawford, Vincent P, 2003, Lying for strategic advantage: Rational and boundedly rational misrepresentation of intentions, *American Economic Review*, 93(1), 133–149.
- [10] DePaulo, Bella M., James J. Lindsay, Brian E. Malone, Laura Muhlenbruck, Kelly Charlton, and Harris Cooper, 2003, Cues to deception, *Psychological Bulletin*, 129, 74–118.
- [11] Erard, Brian, 1992, The influence of tax audits on reporting behavior. In: Joel Slemrod (Ed.), *Why people pay taxes: Tax compliance and enforcement*, pp. 95–114, Ann Arbor (Michigan): University of Michigan Press.
- [12] Erat, Sanjiv and Uri Gneezy, 2012, White Lies, *Management Science* 58(4), 723–733.
- [13] Fischbacher, Urs and Franziska Föllmi-Heusi, 2013, Lies in disguise—An experimental study on cheating, *Journal of the European Economic Association*, 11, 525–547.
- [14] Gibson, Rajna, Carmen Tanner, and Alexander F. Wagner, 2013, Preferences for truthfulness: Heterogeneity among and within individuals, *American Economic Review*, 103, 532–548.
- [15] Gino, Francesca, and Dan Ariely, 2012, The dark side of creativity: original thinkers can be more dishonest, *Journal of personality and social psychology*, 102, 445–459.

- [16] Gino, Francesca, Erin L. Krupka, and Roberto A. Weber, 2013, License to cheat: Voluntary regulation and ethical behavior, *Management Science*, 59, 2187–2203.
- [17] Gneezy, Uri, 2005, Deception: the role of consequences, *American Economic Review*, 95(1), 384–394.
- [18] Greiner, Ben, 2003, The online recruitment system ORSEE: a guide for the organization of experiments in economics. Max-Planck-Institute for Research into Economic Systems, Strategic Interaction Group.
- [19] Guala, Francesco and Luigi Mittone, 2005, Experiments in economics: External validity and the robustness of phenomena, *Journal of Economic Methodology*, 12(4), 495–515.
- [20] Hartwig, Maria and Charles F. Bond, 2011, Why do lie-catchers fail? A lens model meta-analysis of human lie judgments, *Psychological Bulletin*, 137(4), 643–659.
- [21] Harvard Business Review Analytic Services, 2009, Managing across Distance in Today’s Economic Climate: The Value of Face-to-Face Communication, *Research Report*. Available at https://hbr.org/resources/pdfs/comm/british-airways/hbras_ba_report_web.pdf.
- [22] Holm, Håkan J., and Toshiji Kawagoe, 2010, Face-to-face lying – An experimental study in Sweden and Japan, *Journal of Economic Psychology*, 31(3), 310–321.
- [23] Kartik, Navin, 2009, Strategic communication with lying costs, *Review of Economic Studies*, 76, 1359–1395.
- [24] Kartik, Navin, Olivier Tercieux, and Richard Holden, 2014, Simple mechanisms and preferences for honesty, *Games and Economic Behavior*, 83, 284–290.
- [25] Kastlunger, Barbara, Erich Kirchler, Luigi Mittone, and Julia Pitters, 2009, Sequences of audits, tax compliance, and taxpaying strategies, *Journal of Economic Psychology*, 30, 405–418.
- [26] Kleven, Henrik, Martin Knudsen, Claus Kreiner, Soren Pedersen, and Emmanuel Saez, 2011, Unwilling or unable to cheat? Evidence from a tax audit experiment in Denmark, *Econometrica*, 79, 651–692.
- [27] Konrad, Kai A., Tim Lohse, and Salmai Qari, 2016, Compliance with endogenous audit probabilities, forthcoming in: *Scandinavian Journal of Economics*.
- [28] Konrad, Kai A., Tim Lohse, and Salmai Qari, 2014, Deception choice and self-selection – The importance of being earnest, *Journal of Economic Behavior & Organization*, 107, 25–39.
- [29] Lundquist, Tobias, Tore Ellingsen, Erik Gribbe, and Magnus Johannesson, 2009, The aversion to lying, *Journal of Economic Behavior & Organization*, 70(1-2), 81–92.
- [30] Maciejovsky, Boris, Erich Kirchler, and Herbert Schwarzenberger, 2007, Misperception of chance and loss repair: On the dynamics of tax compliance. *Journal of Economic Psychology*, 28(6), 678–691.

- [31] Rezlescu, Constantin, Brad Duchaine, Christopher Y. Olivola, and Nick Chater, 2012, Unfakeable facial configurations affect strategic choices in trust games with or without information about past behavior. *PLoS-ONE*, 7, e34293.
- [32] Sánchez-Pagés, Santiago and Marc Vorsatz, 2007, An experimental study of truth-telling in a sender-receiver game, *Games and Economic Behavior*, 61, 86–112.
- [33] Serra-Garcia, Marta, Eric van Damme, and Jan Potters, 2011, Hiding an inconvenient truth: Lies and vagueness, *Games and Economic Behavior*, 73(1), 244–261.
- [34] Slemrod, Joel and Caroline E. Weber, 2012, Evidence of the invisible: toward a credibility revolution in the empirical analysis of tax evasion and the informal economy, *International Tax and Public Finance*, 19, 25–53.
- [35] Smith, Vernon L., 1994, Economics in the laboratory, *Journal of Economic Perspectives*, 8, 113–131.
- [36] Sutter, Matthias, 2009, Deception through telling the truth?! Experimental evidence from individuals and teams, *Economic Journal*, 119, 47–60.
- [37] Van Swol, Lyn M., Deepak Malhotra, and Michael T. Braun, 2012, Deception and its detection: Effects of monetary incentives and personal relationship history, *Communication Research*, 39(2), 217–238.
- [38] Vanberg, Christoph, 2008, Why do people keep their promises? An experimental test of two explanations, *Econometrica*, 76(6), 1467–1480.
- [39] Vrij, Albert, 2008, *Detecting Lies and Deceit*, 2nd ed. Chichester: Wiley.