Representative evidence on lying costs ☆

Johannes Abeler [a],*, Anke Becker [b], Armin Falk [b]

[a] University of Oxford, United Kingdom
[b] University of Bonn, Germany

ABSTRACT

A central assumption in economics is that people misreport their private information if this is to their material benefit. Several recent models depart from this assumption and posit that some people do not lie or at least do not lie maximally. These models invoke many different underlying motives including intrinsic lying costs, altruism, efficiency concerns, or conditional cooperation. To provide an empirically-validated microfoundation for these models, it is crucial to understand the relevance of the different potential motives. We measure the extent of lying costs among a representative sample of the German population by calling them at home. In our setup, participants have a clear monetary incentive to misreport, misreporting cannot be detected, reputational concerns are negligible and altruism, efficiency concerns or conditional cooperation cannot play a role. Yet, we find that aggregate reporting behavior is close to the expected truthful distribution suggesting that lying costs are large and widespread. Further lab experiments show that this result is not driven by the mode of communication.

© 2014 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/3.0/).

1. Introduction

Situations with asymmetric information are ubiquitous. Most of economic theory assumes that people misreport their private information if this is to their material benefit; behavior is only determined by the trade-off between financial gains from misreporting and monetary fines when misreporting is detected.[1] In contrast, many recent models in various domains of Public Economics (and in Economics more generally) rely on the assumption that people can experience a psychological disutility which holds them back from misreporting, at least to some extent. These models invoke different underlying motives. Kartik et al. (2014), for instance, assume that people face an intrinsic lying cost

and show that in this case the social planner can fully implement a much wider range of social choice rules compared to the standard Maskin (1977) case without lying costs (see, e.g., Matsushima (2008) and Dutta and Sen (2011) for similar assumptions). Many studies about incentive systems for doctors assume that doctors are altruistic towards their patients and thus do not always state the profit-maximizing diagnosis but rather treat patients honestly (e.g., Ellis and McGuire, 1986; Chalkley and Malcomson, 1998). The large literature on "tax morale" (e.g., Lewis, 1982; Cowell, 1990; Andreoni et al., 1998; Slemrod, 2007; Torgler, 2007) demonstrates that many tax payers misreport their income only a little bit or not at all. This literature is usually agnostic about the exact underlying motives but some studies cite efficiency concerns (e.g., Alm et al., 1992), patriotism (Konrad and Qari, 2012), religiosity (Torgler, 2006), fairness (Bordignon, 1993), conditional cooperation (Traxler, 2010) or honesty (Erard and Feinstein, 1994).

To further improve these models and to provide an empirically-validated microfoundation, it is crucial to understand the relevance of the different potential motives. Additionally, understanding these motives could inform the design of more psychologically-realistic policies, e.g., in the area of tax enforcement, that have a higher potential of being successful. In this paper, we focus on intrinsic lying costs and investigate how widespread and how large lying costs are. The ideal data set to answer these questions would allow studying lying costs for a representative sample of the population and in an environment without the confounding effects of strategic interaction (including the

[1] See, e.g., Allingham and Sandmo (1972) on tax evasion, Falkinger (1991) on public good provision, Pitchik and Schotter (1987) on credence goods, along with the seminal Becker (1968) on crime.

levy of fines), reputational or efficiency concerns, or altruism. So far, the best evidence on lying costs comes from experiments conducted in tightly controlled laboratory situations. A robust result is that many subjects misreport their private information to their own advantage but that a substantial share of subjects refrains from reporting the payoff-maximizing type and that some are fully honest (e.g., Gneezy, 2005; Charness and Dufwenberg, 2006; Fischbacher and Föllmi-Heusi, forthcoming; de Haan et al., 2011; Houser et al., 2012; Shalvi et al., 2011; Wibral et al., 2012; Serra-Garcia et al., 2013). These studies are a strong first indicator that lying costs influence behavior. However, lab experiments do not allow for inferences with respect to the prevalence of lying costs in the overall population since they have been conducted almost exclusively with student samples (DellaVigna, 2009; Falk and Heckman, 2009). Also, decision making took place in an austere laboratory environment which might trigger behavior representative only of certain non-lab situations. It could thus be that there are systematic differences between behavior of students in the laboratory and behavior of non-student subjects outside the lab.

To circumvent these limitations, we measure how people report their private information outside the laboratory by calling participants on the phone at their home. Participants were drawn randomly from the German population, yielding a representative sample. An incentivized experiment was embedded in the interview. The experimental setup is related to the design of Fischbacher and Föllmi-Heusi (forthcoming) and is extremely simple: participants were asked to toss a coin and report their type, i.e., either "heads" or "tails". Reporting tails yielded a payoff of 15 euros, which participants could choose to receive in cash or as an Amazon gift certificate, while reporting heads yielded a payoff of zero. Participants thus had a clear monetary incentive to report tails regardless of their true type. It was obvious that the true outcome was only known to the participants, as they tossed the coin privately at home. In this setup, we cannot draw reliable conclusions about the truthfulness of any individual report. But we can learn about aggregate behavior by comparing the distribution of reports to the true distribution of a fair coin (50% tails) and to the payoff-maximizing distribution (100% tails). This indirect observation therefore allows us to study the behavior of subjects in a situation in which private information is kept truly private and in which subjects do not face any risk of detection.[2] Moreover, the decision is non-strategic; altruism does not play a role as the money is not taken from any individual person; and reputational concerns are minimized since the interviewer is a stranger with whom no future interaction can be expected.

If all our participants were rational money maximizers, we would expect that all of them reported tails. If behavior on the phone was similar to previous comparable laboratory experiments (e.g., Houser et al., 2012), we would expect about 75% of subjects reporting tails.

In contrast to these predictions, observed behavior does not statistically differ from everybody reporting honestly. If anything, participants report the payoff maximizing outcome less often than expected under truthful reporting. This latter effect, however, is small and disappears in a second treatment in which participants were asked to report the total number of tails in four consecutive coin tosses and received 5 euros times the number of reported tails. The resulting distribution

of reports in the 4-Coin Treatment is indistinguishable from the distribution under complete truth-telling. Moreover, while previous studies (e.g., Dreber and Johannesson 2008) have found correlations between individual characteristics, like gender, and truth-telling, we do not find any robust correlations between individual characteristics and reporting behavior. This is not surprising if almost all participants report truthfully. Reports are solely determined by chance, namely the coin toss, which cannot be related to any individual characteristic. Our results thus show that lying costs are pervasive and are influencing behavior regardless of gender, religious beliefs, education, or age.

We complement our telephone study with two additional control treatments in the laboratory to better understand what shapes lying costs, in particular the effect of the mode of communication. In both lab treatments subjects reported the outcomes of four consecutive coin tosses. Incentives were the same as in the 4-Coin Treatment in the telephone study: 5 euros times the numbers of tails reported. In the first lab treatment, subjects had to report the outcome directly to an interviewer via the phone, mirroring our telephone study. We observe the same pattern of behavior as in previous lab experiment: subjects lie much more than in the telephone study. In the second control treatment, subjects reported the outcomes by clicking a number between 0 and 4 on the computer screen as in most previous lab experiments. We find that subjects who enter their report by clicking report slightly higher numbers but this difference is not statistically significant. The difference to the telephone study persists: the average report in each lab treatment is higher than in the telephone study. This shows that the mode of communication does not systematically influence reporting behavior strongly and is not driving the widespread truth-telling in our telephone study. We also elicit beliefs about the behavior of other participants and find in all four treatments that participants believe others to lie more than they actually do. Older participants (correctly) believe that lying is less prevalent. In the lab, higher beliefs are correlated with higher own reports. We find no evidence that being a student has a significant impact on behavior, or that the perceived time pressure on the telephone or the limited experience of the survey participants with the abstract design of economics experiments played a role.

Our paper adds to the nascent literature studying lying outside the lab. Previous studies focused on particular groups: Bucciol and Piovesan (2011) study a sample of children and find that many of them lie, unless they are reminded to be honest; Cohn et al. (2013) study prisoners and find that they become less honest when reminded of their criminal identity; and Utikal and Fischbacher (2013) ask a small sample of nuns to report the roll of a dice and find significant downward lying. Studies looking at unethical behavior in less abstract environments include Azar et al. (2013) who find that the majority of customers in a restaurant do not return excessive change. Similarly, Bucciol et al. (2013) study free-riding in public transportation in Italy and find that 43% of passengers evade the fare. We add two features: we study a representative sample and we can investigate the underlying motives by conducting additional lab experiments using the same well-defined decision.

Taken together, our results strengthen the doubts that previous lab experiments have cast on the assumption of zero lying costs: we find evidence for even higher lying costs in the telephone study. This suggests that studying the theoretical implications of such costs (e.g., Kartik et al., 2007, 2014; Doerrenberg et al., 2013) is a promising research avenue. At the same time, it is very likely that altruism, efficiency concerns, etc. are also important factors in the decision to pay taxes or how to treat patients, for example. Future research would need to investigate the relative importance of different motives that hold people back from misreporting and the interactions between motives. Our results also do not mean that lab experiments are uninformative about non-laboratory settings. However, the difference in behavior between our telephone study and our previous lab experiments rather shows how malleable reporting behavior can be. This opens many new questions

---

[2] In other studies concerning how people report their private information (e.g., Gneezy, 2005; Charness and Dufwenberg, 2006), the experimenter knows or will later know the subject's true type (and the subject is aware of this) and can thus judge whether an individual was honest or not. In our experiment, only the participant knows his or her private information. Our setup is thus closer to situations in which information is truly private and only known by the individual, while Gneezy's and Charness & Dufwenberg's setup is more representative of situations in which the private information is known by more than one person, e.g., when filing a joint tax declaration. These papers are also interested in the interaction between sender and receiver, from which we abstract. (See, however, the recent paper by Deck et al. (2013) who do not find an additional effect of promises on cooperation in single-blind and double-blind conditions.)

about how exactly reporting private information depends on the decision-making context. Intuitively, different norms might apply when making such a decision at home, representing a private and familiar environment. Similarly, people could be more attentive to their own moral rules, e.g., abstaining from lying when at home.[3] Irrespective of these differences between lab and field, our study establishes that lying costs are more important than previously assumed and are strongly influencing behavior across different decision environments.

In the next two sections, we present the design of the study and our hypotheses. Section 4 contains the results. We discuss policy implications in Section 5.

2. Design

The computer-assisted telephone interviews were operated by the Institute for Applied Social Sciences (infas), a private and well-known German research institute. They were conducted between November 2010 and February 2011.[4] The average interview lasted 20 min (standard deviation: 5.5 min). Telephone numbers were selected using a random digit dialing technique: numbers were generated randomly based on a data set of all potential telephone numbers in Germany. Only landline numbers were used in this study, as 92.7% of German households have a landline number (Destatis, 2012). The selection of the participant within each household was also random: only the member of the household whose birthday was the most recent among all household members was eligible to participate. We restricted participation to those aged between 18 and 70 years at the time of the interview.[5]

The survey was split into two parts. The first part of the questionnaire consisted of questions relating to the participants' socio-demographic background and their risk and trust preferences. Risk and trust preferences were measured by using subjective self-assessments, using the general risk question of the GSOEP ("How do you consider yourself? Are you in general a rather risk-loving person, or do you try to avoid risks? Use a scale from 1, meaning that you are not at all willing to take risks, to 7, meaning that you are absolutely willing to take risks." (Dohmen et al., 2011)) and the World Value Survey trust question ("Generally speaking: Do you think one can trust other people, or that one should rather be careful when dealing with other people? Please indicate your answer on a scale from 1 to 7, with 1 meaning that one should be careful when dealing with other people, and 7 meaning that one can trust other people."). After this part, the experiment described below took place. After the experiment, participants were asked about their political preferences, their current living and financial situation, their religious beliefs, and their attitudes towards opportunistic behavior and everyday crime. At the very end of the interview, participants were asked to state their belief about other participants' behavior in the experiment.

Before the experiment started the participant was reminded that the resulting data would be anonymized, and that infas and the University of Bonn guaranteed the correct payment. The interviewer then asked the participant to take a coin and explained the rules of the experiment: the task was to toss the coin and report whether heads or tails came up.[6] If the participant reported heads, they received no payment. If the participant reported tails, they would receive 15 euros. Then, the participant was asked to toss the coin and report the outcome. We will call this treatment "1-Coin-Telephone." 658 people participated in this version of our experiment. A translation of the exact experimental instructions can be found in Online Appendix A.

In a second treatment, 94 people were interviewed and participated in the following variation of the experiment. Participants were asked to take a coin, toss it four times, and report the number of times that tails came up. For each time participants reported tails they received 5 euros. Thus, they could earn 0, 5, 10, 15, or 20 euros. We will call this treatment "4-Coin-Telephone." Payment in both treatments could be received either in cash via regular mail or as an Amazon gift certificate code. The alphanumeric 14-digit gift certificate code was transmitted via email or directly on the phone at the end of the interview.

In order to further investigate what influences behavior in the telephone study, in particular the mode of reporting, we additionally conducted two versions of the 4-Coin Treatment in the laboratory. Subjects were students of the University of Bonn studying different majors except Economics. They were seated at a desk with a computer in separate room-high cubicles closed off by curtains. As the experiment took only a few minutes, it was run at the end of the sessions of a different experiment (similar to Fischbacher and Föllmi-Heusi, forthcoming). In the preceding experiment subjects made abstract consumption or labor supply choices which involved no private information and no interaction with other subjects. When the experiment started, subjects were asked to take a coin that was placed in their cubicle, toss it four times, and report how often tails came up. For each time they reported that tails came up they received 5 euros, i.e., up to 20 euros, just like in 4-Coin-Telephone. Their earnings were paid in cash directly after the experiment.[7]

The only difference between the two lab treatments was how the reporting was done. In the first treatment, subjects had to state their report directly to an interviewer via the phone, mirroring our telephone study. After tossing the coin in their cubicle, they were asked to go one-by-one to an adjacent room and pick up the telephone that we had placed there. An interviewer on the other side of the line (whom subjects never met directly) would then ask for their experimental ID and the number of times the coin showed tails. We made sure that other subjects could not hear the conversation. The starting times for the coin tossing was staggered, such that subjects did not have to wait between coin-tossing and reporting. 170 subjects participated in this treatment which we will call "4-Coin-Lab-Tel." This treatment serves to replicate our telephone study as closely as possible in the laboratory. In the second treatment, subjects reported their outcome by clicking a number 0 to 4 on the computer screen, similar to previous lab experiments. 180 subjects participated in the second treatment which we will call "4-Coin-Lab-Click". This treatment serves to investigate whether the mode of communication, i.e., clicking on a computer screen versus reporting to a person via the telephone, influences reporting behavior.

---

[3] Previous research comparing behavior of student samples vs. non-students samples and behavior in the lab vs. outside the lab has in most cases shown little differences (with a few notable exceptions, e.g., Stoop et al., 2012). The strong difference in behavior between our field and lab studies suggests that truth-telling is more context dependent than other behaviors, like cooperation, altruistic behavior, or consumption choices (Abeler and Marklein, 2010). For an overview and critical discussion, see Falk and Heckman (2009), Camerer (2011), or Coppock and Green (2013).

[4] The interviews were conducted in the infas telephone studio. Infas ensures a high quality of interviews by supervising interviews randomly. Supervisors are present in the telephone studio at all times and interviews can be monitored without the interviewer noticing this.

[5] The majority of non-participation was due to no one answering the phone or people hanging up immediately after hearing that a market research firm called. Of the 738 people who started the questionnaire of the 1-Coin Treatment at all and could condition their participation on the content of questionnaire or experiment, 658 participants (89.1%) completed the entire questionnaire and the experiment. Like in all telephone-based surveys, the resulting sample is therefore representative for the part of the population who was at home at the time of call and was willing to participate.

[6] A referee mentioned that some euro coins were reported to not be fair, in particular when spinning the coin. We don't think this is a concern in our study since we asked participants to flip or toss the coin. See Gelman and Nolan (2002) for an explanation on why it is extremely difficult to bias a coin when flipping it.

[7] The instructions for the lab experiment can be found in Online Appendices B and C. The experiment was conducted using ztree and ORSEE (Fischbacher, 2007; Greiner, 2004).

3. Hypotheses

The standard economic prediction in our setup is straightforward: depending on the treatment, people will report tails one or four times, respectively. This is the payoff maximizing outcome as there are no exogenous costs linked to misreporting, no possibility of detection and no fines. The setup is extremely simple and participants should have no trouble identifying the payoff maximizing choice. Moreover, the setup is highly anonymous, discouraging any reputational concerns because of repeated interaction.

If, however, some participants incur a psychological cost or derive direct disutility from falsely reporting their private information per se we should expect both heads and tails to be reported in the experiment. There are a few recent theoretical papers that assume such a cost. For example, Kartik (2009) and Kartik et al. (2007) build on Crawford and Sobel's (1982) cheap-talk model and derive predictions for the case that some agents incur costs when misreporting their private information (see also, e.g., Saran, 2011; Kartik et al., 2014). Assuming some degree of heterogeneity in the incurred costs when misreporting, it is then a question of the trade-off between psychological costs and monetary benefits of misreporting how many participants will report heads and how many report tails.

Participants in 1-Coin-Telephone have to make a clear, binary choice whether to lie or not; if lying costs are related to self-reputation or identity (e.g., Bénabou and Tirole, 2006; Akerlof and Kranton, 2000), lying in such a setting could impact self-reputation or identity more and thus make lying more costly. Participants in 4-Coin-Telephone can make a finer choice between being honest, exaggerating a little bit, or lying maximally; this could render small lies compatible with a positive self-reputation and thus enhance lying (Mazar et al., 2008). Such non-maximal lying has already been shown to be important by Fischbacher and Föllmi-Heusi (forthcoming).

In the telephone study, participants tossed the coin at their home. It was thus obvious that the interviewer could not secretly observe the true outcome of the coin toss.[8] If some participants in our lab experiments (erroneously) believed that the experimenter could observe the true outcome and believed (again erroneously) that misreporting would lead to some negative or unpleasant outcome, we would expect more truth-telling in the laboratory.[9]

Regarding potential differences in reporting behavior according to individual characteristics, we would expect that women are more honest than men (as already shown by Dreber and Johannesson, 2008; Houser et al., 2012). More religious participants would be expected to be more honest, since religious priming leads to less lying and more pro-social behavior (Mazar et al., 2008; Shariff and Norenzayan, 2007). Income could be positively correlated with honesty because of the lower marginal utility of the monetary rewards or negatively correlated because of reverse causality. A similarly ambiguous hypothesis can be derived for education or the social environment, e.g., the size of the community or family status. Along theories of endogenous social

norms (e.g., Traxler 2010; López-Pérez, 2010, 2012), we would expect that higher beliefs about the reporting of other participants are correlated with own high reporting.

4. Results

4.1. Telephone study

Result 1. *In 1-Coin-Telephone, the distribution of actual reports is very close to the truthful distribution; participants report the payoff-maximizing outcome slightly less often than expected if everyone reported truthfully. In 4-Coin-Telephone, the distribution of reports is indistinguishable from the truthful distribution.*

Fig. 1 illustrates aggregate behavior (the dashed line corresponds to the expected distribution if every participant reported the true outcome of the coin toss). 55.6% of participants report heads as the outcome of the coin toss, yielding a payoff of zero, the remaining participants report tails yielding a payoff of 15 euros. The payoff-maximizing outcome is reported slightly less often than in 50% of the cases and although the difference is small in terms of effect size, it is significant (Binomial test, $p = 0.004$) Fig. 2 shows aggregate behavior in 4-Coin-Telephone. Again, reporting behavior follows the expected distribution under complete honesty very closely (the dashed line corresponds to the truthful distribution). In fact, the distribution of reported outcomes is statistically indistinguishable from the truthful distribution (Kolmogorov–Smirnov test, $p = 0.61$; binomial tests of the expected against the observed frequency, all five $p > 0.13$). In particular, and unlike in 1-Coin-Telephone where "too many" people report the payoff-minimizing outcome, there is no significant over-reporting of zero in this treatment.[10] Looking at behavior in both treatments we can therefore summarize that the payoff-maximizing outcome is reported by much fewer participants than expected if no one incurred lying costs. It is also reported less often than suggested by previous lab experimental studies, which find some truth-telling but also many instances of the payoff-maximizing report. Instead, it is close to the distribution that would arise if every participant reported his or her type truthfully.[11]

Previous studies have shown that truth-telling correlates with observable characteristics, e.g. gender or religiosity (Dreber and Johannesson, 2008; Houser et al., 2012; Mazar et al., 2008; Shariff and Norenzayan, 2007). In contrast, if our conjecture that almost all participants report truthfully is correct, an individual's reported outcome will only be driven by their random coin toss; if this is the case, reporting cannot be correlated with any individual characteristic, as these are orthogonal to the

---

[8] We cannot rule out the possibility that, e.g., family members were in the same room with the participant. Behavior, however, does not differ between participants who live alone and those who do not.

[9] Actual anonymity is very high in the telephone study and clearly higher on the telephone than in the lab. Perceived anonymity can and will vary from actual anonymity, for example, participants might believe that someone calling their landline will also know their name or address (which was not the case). However, we don't see a clear reason why perceived anonymity should be higher in the lab than on the phone. The arguments above even suggest that perceived anonymity in the lab is lower than actual anonymity, increasing the telephone-lab difference in perceived anonymity. Either way, there is evidence that the degree of anonymity does not affect behavior much anyway. Fischbacher and Föllmi-Heusi (forthcoming) conduct a double-blind version of their experiment in which both randomization and receiving payment are unobservable by the experimenter. Subjects roll a die in private, take the payment out of an envelope, and then put the envelope back into a box with other envelopes such that it is clear that payments and reports cannot be assigned to any individual. Behavior does not change compared to the baseline treatment, suggesting that (perception of) anonymity plays only a small role.

[10] Note that the sample size in 4-Coin-Telephone is substantially smaller than in 1-Coin-Telephone (94 vs. 658) which reduces the statistical power of the tests. The non-significance is, however, mainly driven by the small effect size. If we (counterfactually) increase the sample size to the usual sample sizes in these kind of experiments (e.g., 389 in the largest treatment in Fischbacher and Föllmi-Heusi (forthcoming) or 251 in Houser et al. (2012) and assume the same shares of reports as in 4-Coin-Telephone, the distribution continues to be indistinguishable from the truthful distribution. This changes only if we increase the sample size beyond 500 (e.g., to 658 as in 1-Coin, Kolmogorov-Smirnov test, $p = 0.02$).

[11] We can only speculate about why some people obviously falsely claim to be of the payoff-minimizing type and why this only happens in 1-Coin-Telephone. The design of the experiment allows ruling out reputational concerns towards the interviewer as an important factor. Privacy concerns could drive this effect: reporting the type that gives zero payoff makes it unnecessary to hand over one's address. The reason why we do not observe such an effect in 4-Coin-Telephone might be that reporting zero to avoid handing over the address was less salient in this treatment. However, we ensured that privacy concerns were minimized in both treatments by giving participants the opportunity to receive the payment as a gift certificate code by email or directly via the phone. 17.2% of eligible participants chose this last payment option which made it unnecessary to hand over any additional contact details. Another possible explanation would be self-image concerns: refraining from easily and safely earning 15 euros could be a strong signal to oneself that one is not greedy and thereby flattering for one's self-image. This interpretation would be in line with how Utikal and Fischbacher (2013) interpret their finding that nuns lie to their monetary disadvantage. We will show more data below which strongly suggests that downward lying is not widespread in our study.
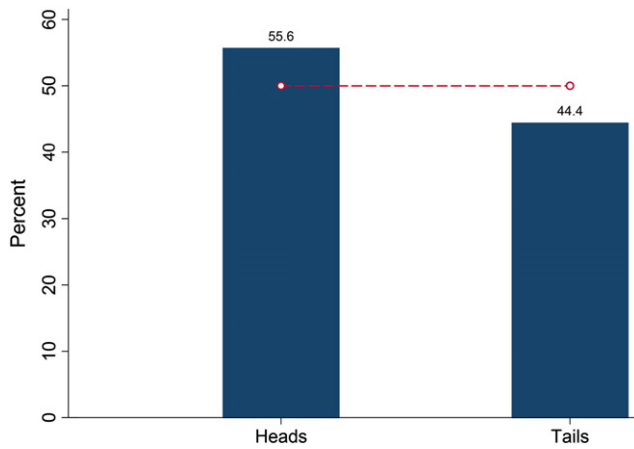
Fig. 1. Aggregate behavior in 1-Coin-Telephone. Reporting heads yielded no payoff; reporting tails yielded a payoff of 15 euros. The dashed line corresponds to the expected distribution if every participant reported the true outcome of their coin toss.

chance move. Therefore, if we do not find such a correlation, our finding of (almost) complete honesty is supported. More specifically, we conduct regression analyses for the two experiments in order to examine whether there are systematic effects of individual characteristics on reporting behavior. First, we regress the report only on clearly exogenous variables such as age and gender, in a second step adding religious denomination. We then include income, the size of the city the individual lives in, and education dummies. Finally, we look at the effect of an individual's religiousness (interacted with denomination), their risk and trust preferences, and their belief about the reporting behavior of other participants.

Result 2. *There is no significant correlation between reporting behavior and any individual characteristic.*

First, we look for potential group differences in terms of reporting behavior in 1-Coin-Telephone by conducting Probit regressions of the reported outcome on the respective characteristics (see Table 2 in Online Appendix E). No characteristic except for one's belief about others' behavior is significantly associated with reporting in the experiment: participants who think many other participants report tails dishonestly, are less likely to report tails themselves. This belief is, however, not significant if we include it as the only explanatory variable ($p = 0.15$). Note in particular that neither gender nor any religion-related variable is significantly correlated with reporting. Conducting the same regressions as in Table 2 using OLS leaves the results unchanged. Next, we
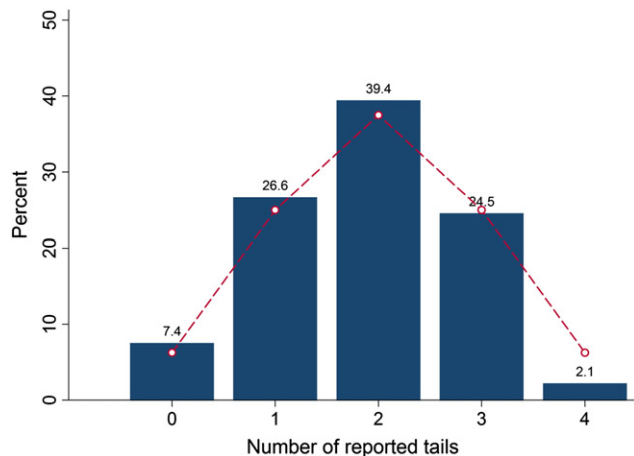

Fig. 2. Aggregate behavior in 4-Coin-Telephone. The payoff was 5 euros times the number of tails reported. The dashed line corresponds to the expected distribution if every participant reported the true outcomes of their coin tosses.

check whether these results also hold in 4-Coin-Telephone. We run Ordered Logit regressions of the reported number of tails on the same explanatory variables as before. Table 3 in the online appendix illustrates the results from this estimation. Only the coefficient for trust is significant. This effect is, however, not robust to the inclusion of other explanatory variables. The effect is also not present in 1-Coin-Telephone. In contrast to 1-Coin-Telephone, the belief coefficient shows no significant association with reporting behavior in this treatment and the point estimate has the opposite sign. We will discuss the data on beliefs in more detail in Section 4.3.

Two further aspects of our analysis are worth noting. First, when running OLS regressions using the same predictor variables as above, we find that only two of the 10 specifications have an adjusted R^2 above 0 (below 0.004), all other adjusted R^2 values are negative. Moreover, the resulting adjusted R^2 tends to decrease in the number of included variables. This again underlines our conclusion: the tested predictor variables do not increase explained variance in the dependent variable compared to pure chance. Second, we also test the correlations between reported number and answers to the survey questions that we did not include in the main specifications of Tables 2 and 3. These include a person's citizenship and country of birth, various personal characteristics, a person's current job or educational situation and their current or recent position in the professional hierarchy, a person's willingness to tell white lies in different situations, a person's family status and living situation (whether one lives with a partner and the number of people belonging to the household), the frequency of church attendance, a person's political preference, and the individual's tendency to behave in an opportunistic way as well as the belief about others' willingness to behave like that. Testing these variables as predictors in Probit and Ordered Logit regressions in the two different data sets, akin to Tables 2 and 3, we find no robust association between any of them and reporting behavior. In particular, this means that students and non-students do not behave differently in our sample. This holds when we consider current students or include former students as well (Kolmogorov–Smirnov tests, all $p > 0.409$). It is thus not a student vs. non-student difference, e.g., a difference in education, age, cognitive skills, or socio-demographic background, which drives the difference between our results and previous lab experiments. Summing up, the overall picture is confirmed: no individual characteristic, whether exogenous or endogenous, is systematically associated with reporting behavior suggesting that almost all participants in our study tell the truth. It could still be that a subgroup of people, which we cannot identify with our background information, reports tails more often than actually true while another subgroup reports tails less often. This could result in the two effects offsetting each other, which would result in a similar picture of aggregate behavior. However, we consider this to not be likely as our analysis shows that this is not the case for any of the numerous subgroups that we can identify with our data. More importantly, such an effect would further need to recreate the distinct distributions of Figs. 1 and 2 which is implausible.

4.2. Laboratory experiment

To further investigate the motivations underlying behavior in the telephone study, we conducted two 4-Coin Treatments as laboratory experiments. We will first discuss the 4-Coin-Lab-Tel treatment which keeps the mode of communication as in the telephone study: subjects had to report their result over the phone directly to an experimenter.[12] Subsequently, we compare this treatment to 4-Coin-Lab-Click, in which subjects reported their number by clicking

---

[12] It was obvious to the subjects that the experimenter on the phone was not the same person as the experimenter in the lab, since the experimenter in the lab coordinated the procedure of calling subjects one-by-one into the separate room with the phone.

a button on a computer screen as in previous lab experiments. This second comparison will allow us to disentangle the influence of the mode of communication.[13]

Result 3. *Subjects in 4-Coin-Lab-Tel report substantially higher numbers than subjects in 4-Coin-Telephone.*

The upper panel of Fig. 3 shows aggregate behavior in 4-Coin-Lab-Tel: most subjects refrain from reporting the maximal outcome, forgoing on average 6.82 euros, quite a considerable amount compared to the average hourly student wage in Germany of about 10 euros. At the same time, behavior is significantly different from the distribution expected under truthful reporting, the dashed line in the figure (Kolmogorov–Smirnov test, $p < 0.001$; binomial tests, all five $p < 0.009$). This replicates previous findings in the lab: many subjects lie but often not maximally. Reporting behavior also deviates strongly from what we have observed in the telephone study: reports are significantly higher in 4-Coin-Lab-Tel than in 4-Coin-Telephone. In Table 1, columns 1 and 2, we regress the reported number of tails on a dummy for being in the lab, a dummy for 4-Coin-Lab-Click and controls for age and gender. The lab dummy is highly significant.[14] We find the same result if we compare 4-Coin-Telephone and 4-Coin-Lab-Tel using a two-sample Kolmogorov–Smirnov test ($p < 0.001$). These results demonstrates that our 4-coin randomization mechanism does not drive the truthful behavior in 4-Coin-Telephone and that, by moving our telephone setup to the laboratory, we are able to strongly change behavior (as we showed within the telephone study, this is not driven by subjects being students per-se). How big is the additional effect if we also change the mode of communication?

Result 4. *Subjects in 4-Coin-Lab-Click report slightly higher numbers than subjects in 4-Coin-Lab-Tel but this difference is not statistically significant. Only the report of 4 occurs significantly more often in 4-Coin-Lab-Click; the reports of 0, 1, 2, and 3 are not different across treatments. Reports in 4-Coin-Lab-Click are significantly higher than in 4-Coin-Telephone.*

The lower panel of Fig. 3 shows aggregate behavior in 4-Coin-Lab-Click. The distribution of reports is very similar to the one in 4-Coin-Lab-Tel, the average report being only slightly higher (2.78 in Click vs. 2.64 in Tel). The overall distribution and the average report are not significantly different across the two treatments (two-sample Kolmogorov–Smirnov test, $p = 0.136$; Ordered Logit in Table 1, columns 1 and 2, both $p > 0.096$). The share of subjects reporting 0, 1, 2 or 3 are also not significantly different (tests of proportion, all $p > 0.100$). However, subjects in 4-Coin-Lab-Click report 4 significantly more often ($p = 0.007$).[15] At the same time, behavior in 4-Coin-Lab-Click is markedly different from 4-Coin-Telephone (two-sample Kolmogorov Smirnov test, $p < 0.001$, Ordered Logit in Table 1, columns 1 and 2, F-test, both $p < 0.001$). Overall, our data thus show that the mode of communication does not have a strong effect on behavior and cannot explain the difference between our telephone study and previous lab experiments. This result is further confirmed by Waubert De Puiseau and Glöckner (2012) who also find considerable truth-telling at home,
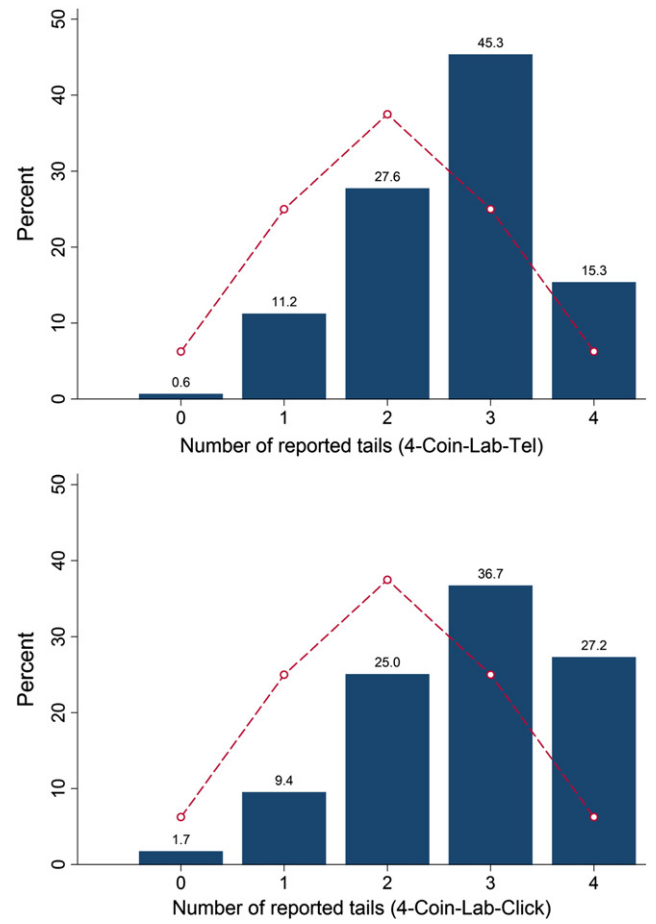

Fig. 3. Aggregate behavior in the 4-Coin-Lab-Tel (upper panel) and 4-Coin-Lab-Click treatments (lower panel). The payoff was 5 euros times the number of tails reported. The dashed line corresponds to the expected distribution if every participant reported the true outcomes of their coin tosses.

though not as extreme as in our data, using an online panel in which participants answered questions at home by clicking on a computer screen. Houser et al. (2012) conduct a 1-coin lab experiment and find similar levels of lying as in our lab experiments, replicating the other side of our results.

One could think that one reason why behavior in the telephone study differs is a perceived time pressure on the telephone which might make lying more difficult. However, we measure response times in the laboratory and do not find a correlation with the report (Ordered Logit, $p = 0.108$).[16] If anything, the report in the lab is higher for short decision times. This mirrors results of Shalvi et al. (2012) who impose exogenous time pressure in a similar lab experiment and who find that subjects become less honest under time pressure. Taken together, these results suggest that behavior in the telephone study is not driven by perceived time pressure. We also find no correlation of the number of previous participations in other lab experiments with reporting behavior in the lab ($p = 0.578$), suggesting that the limited experience participants of the telephone study have with the abstract design of economics experiments does not play a role. It rather seems that different norms apply when making a reporting decision at home,

---

[13] We asked subjects to toss the coin four times instead of only once, to be able to replicate non-maximal overreporting, one of the main results of Fischbacher and Föllmi-Heusi (forthcoming). See Houser et al. (2012) and Bucciol and Piovesan (2011) for studies with a single coin toss; both also find significant overreporting.

[14] We use Ordered Logit regressions in Table 1. All results, including the ones discussed below, also obtain when we use OLS instead.

[15] Two subjects in 4-Coin-Lab-Click told us that they "accidentally clicked the wrong button" and thus wanted to change their report; both subjects wanted to reduce their report, one subject from 4 to 2 and the other from 4 to 0. The data shown here includes their final report as they received this report as payoff. Results stay very similar when we consider their initial click.

[16] We restrict the sample to 4-Coin-Lab-Click as the response time is measured very noisily in 4-Coin-Lab-Tel where we cannot distinguish the actual decision time from the walking to the next room, reporting, and coming back to the cubicle.

Table 1
Comparison of 4-Coin Treatments.

Dependent variable:	Number of reported tails (0–4)				Belief about other participants	
	(1)	(2)	(3)	(4)	(5)	(6)
1 if either lab treatment	1.370***	1.079***	1.242***	1.260***	0.074**	−0.072
	(0.225)	(0.289)	(0.246)	(0.334)	(0.035)	(0.060)
1 if 4-Coin-Lab-Click	0.334*	0.307	0.230	0.204	0.098***	0.100***
	(0.201)	(0.203)	(0.204)	(0.206)	(0.031)	(0.030)
1 if female		−0.345*		−0.371*		−0.017
		(0.183)		(0.190)		(0.028)
Age		−0.013		0.001		−0.006***
		(0.009)		(0.010)		(0.002)
Belief about other participants			2.114***	2.115***		
			(0.356)	(0.358)		
N. Obs.	444	443	425	424	425	424

Notes: Ordered Logit Estimates (columns 1–4) and Tobit estimates (columns 5–6). Robust standard errors are in parentheses. The sample includes all 4-Coin Treatments, i.e., 4-Coin-Telephone, 4-Coin-Lab-Tel, and 4-Coin-Lab-Click. "Belief about other participants" is the belief of this participant about the share of participants who report to have tossed more tails than they actually did and who report 4 tails (see text for details about the question). Significance at the 1, 5, and 10% level is denoted by ***, **, and *, respectively.

representing a private and familiar environment, compared to in the lab where other, more selfish norms might be triggered.[17]

We showed above that women do not report differently from men in the telephone study. As one can see from Table 1, women do report lower numbers in the lab. This effect is only weakly significant in the sample of all three 4-Coin Treatments, i.e., also including 4-Coin-Telephone which dilutes the effect, and becomes significant if we restrict the sample to the two lab treatments ($p = 0.027$ and $p = 0.046$ in regressions akin to columns 2 and 4 of Table 1).

4.3. Beliefs about other participants

Previous studies (e.g., López-Pérez, 2010; Diekmann et al. 2011) have investigated the relationship of reporting behavior and the beliefs about what other people report. Since our telephone and lab settings generate strong differences in reporting behavior, we next examine whether there is a similar difference in beliefs and whether this could help explain the differences in behavior.

In all four treatments, we elicited beliefs about the reporting behavior of the other participants. We will mainly focus on analyzing beliefs in the 4-Coin Treatments as the outcome variable is richer and we have additional treatments. In the 4-Coin Treatments, subjects were asked two questions regarding their beliefs about the behavior of other subjects in their treatment (the question referred to 1000 participants in 4-Coin-Telephone): "We are conducting this experiment also with 100 other participants. How many of these 100 participants do you think report tails more often than they actually tossed?" and "How many of these X overreporting participants do you think report that they tossed tails in each of the four coin tosses?"[18] We will use the answers as direct measure of the belief about the share of liars and about the share of maximal liars. Using a very simple model, we can also combine the two answers to back out the implied belief about the average report in the population or the share of participants reporting the payoff-maximizing outcome. The model assumes, similar to Kartik (2009) and Mazar et al. (2008), that participants expect others to face a psychological lying cost which is increasing and convex in the size of the lie and

which might be heterogeneous between participants. Online Appendix D describes the model and the belief measures in more detail.

Result 5. *In all treatments, participants believe others to overreport more than they actually do.*

Fig. 4 compares average beliefs with average actual behavior for each treatment. We take as variable of interest the share of participants who report the payoff-maximizing outcome, i.e., 4 tails in the 4-Coin Treatments and tails in 1-Coin-Telephone. Since we expected participants to be unfamiliar with the true distribution of the sum of four coin tosses, we didn't ask directly for their belief about this share. We are able to calculate it, given the assumption of convex lying costs, from the two questions for the 4-Coin Treatments: it is the share of liars who report 4 (question 2) plus the share of honest 4's (the probability of a true 4 times (1 − answer 1)). Since we do not observe whether an individual overreports we cannot directly compare the two answers to actual behavior.[19] We find in all four treatments that participants believe that others overreport more than they actually do. The differences are highly significant (t-tests, all $p < 0.001$). The same results obtain when we consider the average reported number as variable of interest.[20]

What shapes these beliefs?

Result 6. *Older participants believe others to overreport less. Participants expect more overreporting when participants can enter their report by clicking on the screen.*

In Table 1, columns 5 and 6, we regress the answer to the second question on treatment dummies, a gender dummy and age (the table only considers the 4-Coin Treatments). We find that subjects in 4-Coin-Lab-Click believe others to overreport more than subjects in 4-Coin-Lab-Tel. Being in the laboratory seems to increase beliefs (column 5) but this effect goes away once we control for age (column 6). Participants in the telephone study are on average much older than the student sample in the lab and older participants expect others to overreport less. This means that the beliefs of older participants in the telephone study are closer to actual behavior than the beliefs of younger participants. The same effect of age is present in 1-Coin-Telephone ($p < 0.001$). Using the answer to the first question, or the belief about the average report

---

[17] Our lab and field experiments differ in a couple of other respects which we cannot disentangle: subjects in the lab, for example, know that other subjects are in the same room, even though they are separated by walls and curtains, while at least some telephone participants will be alone; this might lead to different norms being triggered as suggested above. Furthermore, the telephone survey came as a surprise to participants while subjects in the lab experiment signed up in advance and expected to participate and to earn money. Abeler (2013) explores the interaction of expectations and honesty and suggests that higher expectations could lead to less honesty, in line with our results.
[18] In 1-Coin-Telephone, we only asked one question: "How many of the participants report tails although they tossed heads?"

[19] In Fig. 4 we assume that subjects in 1-Coin-Telephone expect all tossed tails to be reported as tails.
[20] If some participants care about the distribution of behavior among all participants, i.e., a kind of group reputation, the wrong belief could be a potential reason for why we find that some people lie to their monetary disadvantage in 1-Coin-Telephone: their behavior could be motivated by a desire to compensate for others' behavior whom they (falsely) believe to be lying. In 4-Coin-Telephone, such a strategy is not fruitful as too many zeros would not help the group reputation.
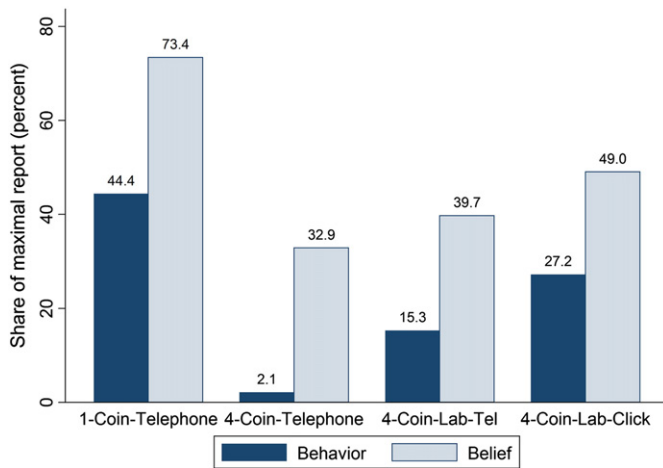
Fig. 4. Share of maximal reports across treatments. The maximal report is 4 in the 4-Coin Treatments and 1 in 1-Coin-Telephone. The dark bars depict actual behavior. The light-colored bars depict the average belief of participants in each treatment about the behavior of the other participants in their treatment.

or the belief about the share of participants reporting 4 does not change any of the results.

Result 7. *In the lab, participants who believe that others report high numbers also report higher numbers themselves.*

We discussed above that there is no robust correlation between reports and beliefs in the telephone study. In Table 1, columns 3 and 4, we study the correlation of reports and beliefs for the 4-Coin Treatments in lab and field. We regress the reported number of coin tosses on treatment dummies, controls for gender and age and on the answer to the second belief question. We find that participants who believe others to report high numbers also report higher numbers themselves. If we exclude 4-Coin-Telephone from the analysis, the coefficient on the belief variable becomes even bigger and stays significant. One could interpret this finding as yet another indication that almost all participants are honest in the telephone study because, if some were not, we should also find a correlation with beliefs in the telephone study (similar to the gender effect we do not find). Furthermore, since beliefs are on average higher in 4-Coin-Tel-Click, the difference between the two lab treatments – which is barely significant in column 1 – becomes even smaller once we control for beliefs. These results are again robust to the exact belief measure we use.

The direction of causality between beliefs and behavior is obviously unclear in our setting. On the one hand, it could be that a high belief induces participants to also report higher numbers. This would be in line with a notion that moral norms are endogenous to the beliefs people hold about the behavior of their peers (see, e.g.,Traxler, 2010; López-Pérez, 2010, 2012). Diekmann et al. (2011) provide causal evidence that higher beliefs lead to higher reports. If this is the mechanism for the correlation between beliefs and behavior, it is even more surprising that participants, in particular in the telephone study, decided to refrain from exploiting the opportunity to receive a considerable amount of money when they believed that many others would do so. On the other hand, the causality might run in the opposite direction if participants ex-post justify their own high report with a stated belief that others also overreport.

5. Conclusion

Using a representative sample of the German population we conducted telephone interviews during which respondents participated in an incentivized experiment. Depending on the treatment, they could earn money by reporting tails as the outcome of one or four coin tosses. We find that almost all participants report their coin toss(es) honestly: the distributions of reports are extremely close to the true distribution of a fair coin toss or four coin tosses, respectively. Moreover, reports are not correlated with any individual characteristic, including gender which has been shown to predict honesty in previous lab studies. We conducted additional laboratory experiments to study the motives underlying the behavior on the phone. While reports are generally higher in the lab than in the telephone study we find little evidence that the mode of communication (reporting directly to someone via the phone vs. clicking a number on a computer) influences behavior. Being a student has also no effect.

Our results underline doubts about the generalizability of economic models which assume that people always lie maximally when it is financially beneficial. Apparently, people do not only care for the trade-off between financial gains from misreporting and the monetary fines when misreporting is detected (cf. Becker, 1968). Our results instead support models like Erard and Feinstein (1994), Kartik et al. (2014) or Doerrenberg et al. (2013) which assume that many people do not lie or do not lie maximally; intrinsic lying costs could be a potential microfoundation for these and similar models. The effect of patriotism and religiosity on tax morale (Konrad and Qari, 2012; Torgler, 2006), for example, could also work through an increased lying cost.

The strong differences we find between telephone and lab environment suggest that lying costs are stronger in our setting outside the lab. It seems that different norms apply when reporting private information at home. Similarly, it might be that the familiar and intimate environment of one's own home reinforces one's personal identity and renders personal moral standards more salient. This is in line with recent evidence by Cohn et al .(2013) who conduct a similar experiment with prisoners. They find that priming prisoners with their criminal identity reduces honesty. Lab experiments, in turn, could be more representative of decisions for which people take on a particular role or identity in addition to their private identity.

At the same time, this study does not imply that everybody always reports their private information truthfully. The level of lying costs seems rather to be influenced by the context in which people are asked to report their type (see also Mazar and Ariely, 2006; Mazar et al., 2008). The difference in behavior on the phone and in the lab shows how malleable reporting behavior can be. Our results therefore point to important policy implications: institutions, e.g., tax authorities, could make use of the context dependence of reporting behavior when designing decision-making environments. As we find strong evidence for widespread lying costs, appropriate mechanisms might be much less complex than those resulting when assuming that agents have no qualms about lying. It might be possible to change reporting behavior in simple and low-cost ways in the spirit of libertarian paternalism (Thaler and Sunstein, 2003). Further research is necessary to uncover what the crucial aspects of the decision-making environment are that induce truth-telling.

Appendix A. Supplementary material

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.jpubeco.2014.01.005.

References

Abeler, J. 2013, 'Preferences for truth-telling', *mimeo*.
Abeler, J., Marklein, F., 2010. Fungibility, Labels, and Consumption. CeDEx Discussion Paper.
Akerlof, G., Kranton, R., 2000. Economics and identity. Q. J. Econ. 3, 715–753.
Allingham, M.G., Sandmo, A., 1972. Income tax evasion: a theoretical analysis. J. Public Econ. 1, 323–338.
Alm, J., McClelland, G.H., Schulze, W.D., 1992. Why do people pay taxes? J. Public Econ. 48 (1), 21–38.
Andreoni, J., Erard, B., Feinstein, J., 1998. Tax compliance. J. Econ. Lit. 36 (2), 818–860.

Azar, O.H., Yosef, S., Bar-Eli, M., 2013. Do customers return excessive change in a restaurant? A field experiment on dishonesty. J. Econ. Behav. Organ. 93, 219–226.

Becker, G.S., 1968. Crime and punishment: an economic approach. J. Polit. Econ. 76 (2), 169–217.

Bénabou, R., Tirole, J., 2006. Incentives and prosocial behavior. Am. Econ. Rev. 96 (5), 1652–1678.

Bordignon, M., 1993. A fairness approach to income tax evasion. J. Public Econ. 52 (3), 345–362.

Bucciol, A., Landini, F., Piovesan, M., 2013. Unethical behavior in the field: demographic characteristics and beliefs of the cheater. J. Econ. Behav. Organ. 93, 248–257.

Bucciol, A., Piovesan, M., 2011. Luck or cheating? A field experiment on honesty with children. J. Econ. Psychol. 32 (1), 73–78.

Camerer, C., 2011. The Promise and Success of Lab-Field Generalizability in Experimental Economics. SSRN Discussion Paper.

Chalkley, M., Malcomson, J.M., 1998. Contracting for health services when patient demand does not reflect quality. J. Health Econ. 17, 1–19.

Charness, G., Dufwenberg, M., 2006. Promises and partnership. Econometrica 74 (6), 1579–1601.

Cohn, A., Maréchal, M., Noll, T. 2013, 'Saliency of criminal identity causes dishonest behavior: An experiment behind bars', mimeo.

Coppock, A. Green, D. 2013, 'Assessing the correspondence between experimental results obtained in the lab and field: A review of recent social science', mimeo.

Cowell, F.A., 1990. Cheating the Government. MIT Press, Cambridge, Mass.

Crawford, V.P., Sobel, J., 1982. Strategic information transmission. Econometrica 50 (6), 1431–1451.

de Haan, T., Offerman, T., Sloof, R., 2011. Money Talks? An Experimental Investigation of Cheap Talk and Burned Money. Discussion Paper.

Deck, C., Servátka, M., Tucker, S., 2013. An examination of the effect of messages on cooperation under double-blind and single-blind payoff-procedures. Exp. Econ. 16 (4), 597–607.

DellaVigna, S., 2009. Psychology and economics: evidence from the field. J. Econ. Lit. 47 (2), 315–372.

Destatis, 2012. Federal Statistical Office of Germany. Annual Report.

Diekmann, A., Przepiorka, W., Rauhut, H., 2011. Lifting the Veil of Ignorance: An Experiment on the Contagiousness of Norm Violations. CESS Discussion Paper.

Doerrenberg, P., Duncan, D., Fuest, C., Peichl, A., 2013. Nice Guys Finish Last: Are People with Higher Tax Morale Taxed More Heavily? CESifo Discussion Paper.

Dohmen, T., Falk, A., Huffman, D., Sunde, U., Schupp, J., Wagner, G.G., 2011. Individual risk attitudes: measurement, determinants, and behavioral consequences. J. Eur. Econ. Assoc. 9 (3), 522–550.

Dreber, A., Johannesson, M., 2008. Gender differences in deception. Econ. Lett. 99 (1), 197–199.

Dutta, B., Sen, A., 2011. Nash implementation with partially honest individuals. Games Econ. Behav. 74 (1), 154–169.

Ellis, R.P., McGuire, T.G., 1986. Provider behavior under prospective reimbursement — cost sharing and supply. J. Health Econ. 129–151.

Erard, B., Feinstein, J.S., 1994. Honesty and evasion in the tax compliance game. RAND J. Econ. 25 (1), 1–19.

Falk, A., Heckman, J.J., 2009. Lab experiments are a major source of knowledge in the social sciences. Science 326 (5952), 535–538.

Falkinger, J., 1991. On optimal public good provision with tax evasion. J. Public Econ. 45 (1), 127–133.

Fischbacher, U., 2007. z-Tree: Zurich toolbox for ready-made economic experiments. Exp. Econ. 10 (2), 171–178.

Fischbacher, U., Föllmi-Heusi, F., 2014. Lies in disguise — an experimental study on cheating. J. Eur. Econ. Assoc. (forthcoming).

Gelman, A., Nolan, D., 2002. You can load a die, but you can't bias a coin. Am. Stat. 56 (4), 308–311.

Gneezy, U., 2005. Deception: the role of consequences. Am. Econ. Rev. 95 (1), 384–394.

Greiner, B., 2004. An online recruitment system for economic experiments. Res. Sci. Comput. 63, 79–93.

Houser, D., Vetter, S., Winter, J.K., 2012. Fairness and cheating. Eur. Econ. Rev. 56, 1645–1655.

Kartik, N., 2009. Strategic communication with lying costs. Rev. Econ. Stud. 76 (4), 1359–1395.

Kartik, N., Ottaviani, M., Squintani, F., 2007. Credulity, lies, and costly talk. J. Econ. Theory 134 (1), 93–116.

Kartik, N., Tercieux, O., Holden, R., 2014. Simple mechanisms and preferences for honesty. Games Econ. Behav. 83, 284–290.

Konrad, K.A., Qari, S., 2012. The last refuge of a scoundrel? Patriotism and tax compliance. Economica 79 (315), 516–533.

Lewis, A., 1982. The Psychology of Taxation. Martin Roberson, Oxford.

López-Pérez, R., 2010. Guilt and shame: an axiomatic analysis. Theor. Decis. 69, 569–586.

López-Pérez, R., 2012. The power of words: a model of honesty and fairness. J. Econ. Psychol. 33 (3), 642–658.

Maskin, E.S., 1977. Nash Equilibrium and Welfare Optimality. MIT Discussion Paper.

Matsushima, H., 2008. Role of honesty in full implementation. J. Econ. Theory 139 (1), 353–359.

Mazar, N., Amir, O., Ariely, D., 2008. The dishonesty of honest people: a theory of self-concept maintenance. J. Mark. Res. 45 (6), 633–644.

Mazar, N., Ariely, D., 2006. Dishonesty in everyday life and its policy implications. J. Public Policy Mark. 25 (1), 117–126.

Pitchik, C., Schotter, A., 1987. Honesty in a model of strategic information transmission. Am. Econ. Rev. 77 (5), 1032–1036.

Saran, R., 2011. Bilateral trading with naive traders. Games Econ. Behav. 72 (2), 544–557.

Serra-Garcia, M., van Damme, E., Potters, J., 2013. Lying about what you know or about what you do. J. Eur. Econ. Assoc. 11 (5), 1204–1229.

Shalvi, S., Dana, J., Handgraaf, M.J., De Dreu, C.K., 2011. Justified ethicality: observing desired counterfactuals modifies ethical perceptions and behavior. Organ. Behav. Hum. Decis. Process. 115 (2), 181–190.

Shalvi, S., Eldar, O., Bereby-Meyer, Y., 2012. Honesty requires time (and lack of justifications). Psychol. Sci. 23 (10), 1264–1270.

Shariff, A.F., Norenzayan, A., 2007. God is watching you: priming god concepts increases prosocial behavior in an anonymous economic game. Psychol. Sci. 18 (9), 803–809.

Slemrod, J., 2007. Cheating ourselves: the economics of tax evasion. J. Econ. Perspect. 1, 25–48.

Stoop, J., Noussair, C.N., Van Soest, D., 2012. From the lab to the field: cooperation among fishermen. J. Polit. Econ. 120 (6), 1027–1056.

Thaler, R.H., Sunstein, C.R., 2003. Libertarian paternalism. Am. Econ. Rev. 93 (2), 175–179.

Torgler, B., 2006. The importance of faith: tax morale and religiosity. J. Econ. Behav. Organ. 61 (1), 81–109.

Torgler, B., 2007. Tax Compliance and Tax Morale. Edward Elgar Publishing, Cheltenham.

Traxler, C., 2010. Social norms and conditional cooperative taxpayers. Eur. J. Polit. Econ. 26, 89–103.

Utikal, V., Fischbacher, U., 2013. Disadvantageous lies in individual decisions. J. Econ. Behav. Organ. 85, 108–111.

Waubert De Puiseau, B., Glöckner, A. 2012, 'Investigating cheating behavior in students compared to the general public', mimeo.

Wibral, M., Dohmen, T., Klingmüller, D., Weber, B., Falk, A., 2012. Testosterone administration reduces lying in men. PLoS One 7 (10), e46774.