Karl-Franzens-University Graz
Faculty of Social and Economic Sciences

To claim or not to claim: Anonymity, reciprocal externalities and honesty

Christian Schitter, Jürgen Fleiß, Stefan Palan

  

Working Paper 2017-01

May 24, 2017

Subsequently published as:

Schitter, C., Fleiß, J., Palan, S., "To claim or not to claim: Anonymity, reciprocal externalities and honesty", Journal of Financial Markets 71, 2018, 13-63, DOI: 10.1016/j.joep.2018.09.006.

An electronic version of the paper may be downloaded from:

University of Graz: sowi.uni-graz.at/forschung/working-paper-series/
RePEc: ideas.repec.org/s/grz/wpsses.html

Working Paper Series
Faculty of Social and Economic Sciences
Karl-Franzens-University Graz
ISSN 2304-7658
sowi.uni-graz.at/forschung/working-paper-series/
sowi-wp@uni-graz.at

UNIVERSITY OF GRAZ

To claim or not to claim: Anonymity, reciprocal externalities and honesty

Christian Schitter[a], Jürgen Fleiß[b], Stefan Palan[a,c,]*

Working Paper 2017-01

May 24, 2017

Abstract

This paper investigates the determinants of (dis)honesty of reporters filing unverified claims for money. First, does honest reporting increase when each reporter's unverified claim is made public? We present experimental evidence to this effect. The driver behind this is activation of the preference for appearing honest. Second, does honest reporting increase when it is public knowledge that reporters' claims affect others and reporters are reciprocally affected by others' claims? We find no such effect. Fear of losing out against others who untruthfully claim too much may outweigh honesty and pro-social considerations.

Keywords: honesty, anonymity, externalities, shame, guilt, pro-social preferences

Subsequently published as:

Schitter, C., Fleiß, J., Palan, S., "To claim or not to claim: Anonymity, reciprocal externalities and honesty", Journal of Financial Markets 71, 2018, 13-63, DOI: 10.1016/j.joep.2018.09.006.

[a] University of Graz, Department of Banking and Finance, Universitaetsstrasse 15, 8010 Graz, AUSTRIA

[b] University of Graz, Department of Corporate Leadership and Entrepreneurship, Elisabethstraße 50b, 8010 Graz, AUSTRIA

[c] University of Innsbruck, Department of Banking and Finance, Universitaetsstrasse 15, 6020 Innsbruck, AUSTRIA

[*] Corresponding author. Tel.: +43(316)380-7306, E-Mail stefan.palan@uni-graz.at.

Any opinions expressed herein are those of the author(s) and not those of the Faculty of Social and Economic Sciences. Working Papers receive only limited review and often represent preliminary work. They are circulated to encourage discussion. Citation of such a paper should account for its provisional character. A revised version may be available directly from the author.

Working Paper Series
Faculty of Social and Economic Sciences
Karl-Franzens-University Graz
ISSN 2304-7658
sowi.uni-graz.at/forschung/working-paper-series/
sowi-wp@uni-graz.at

UNIVERSITY OF GRAZ

To claim or not to claim: Anonymity, reciprocal externalities and honesty

Christian Schitter[a,*], Jürgen Fleiß[b], Stefan Palan[a,c]

[a]*Department of Banking and Finance, University of Graz*
[b]*Department of Corporate Leadership and Entrepreneurship, University of Graz*
[c]*Department of Banking and Finance, University of Innsbruck*

Abstract

This paper investigates the determinants of (dis)honesty of reporters filing unverified claims for money. First, does honest reporting increase when each reporter's unverified claim is made public? We present experimental evidence to this effect. The driver behind this is activation of the preference for appearing honest. Second, does honest reporting increase when it is public knowledge that reporters' claims affect others and reporters are reciprocally affected by others' claims? We find no such effect. Fear of losing out against others who untruthfully claim too much may outweigh honesty and pro-social considerations.

Keywords: Honesty, anonymity, externalities, shame, guilt, pro-social preferences

1. Introduction

In many economic transactions, the possibility of dishonesty poses a significant problem. This is especially salient in the presence of information asymmetry where one party has to rely on unverified information reported by the other. In cases like income tax reports or social insurance claims, the verification of the information provided can drain considerable resources. For other cases, like damage size claims in theft insurance, self-reports are almost completely unverifiable. Insurance companies often have to rely on measures other than verification of the report to increase claim honesty under such circumstances. It is, therefore, unsurprising that both behavioral economists and psychologists have dedicated considerable attention to factors fostering and impeding honesty in recent years (e.g., Rosenbaum et al., 2014; Irlenbusch and Villeval, 2015; Jacobsen et al., 2017).

The situations described before have in common that reporters usually deal with anonymous institutions where many mechanisms fostering honesty in personal interactions may not apply. Tax authorities and insurance companies act as intermediaries to collect resources for, or distribute resources from, a common pool, jointly owned by a large group of people.

*Corresponding author

Email addresses: `christian.schitter@edu.uni-graz.at` (Christian Schitter), `juergen.fleiss@uni-graz.at` (Jürgen Fleiß), `stefan.palan@uni-graz.at` (Stefan Palan)

Preprint submitted to Faculty of Social and Economic Sciences Graz Working Paper Series May 24, 2017

This leads to an increased perceived distance from the people who are hurt by reporters' dishonesty. The perceived distance may both reduce the reporters' feelings of being observed and may disguise the fact that dishonest reporting affects other people.

Several institutions and economic actors resort to a reduction in the anonymity of reporters and to providing information on externalities created by untruthful reports to increase compliance. Micro-insurance programs, for example, in developing countries and, more recently, peer-to-peer (P2P) insurance schemes in western countries[1] create small risk pools. In such micro-collectives, insurance holders learn both the size of each reported damage and who made the claim for compensation. Members of the collective are thus aware of how the pool is affected by each individual claim. In the insurance literature, the assumed effectiveness of this approach is attributed to a higher degree of identifiability of victims (Köneke et al., 2015, chapter 15.12). Biener et al. (2016) study the effectiveness of having joint liability as a group and find that such risk pooling can increase effort and reduce moral hazard in micro-insurance schemes. However, they do not address opportunistic fraud under such mechanisms. Similarly, in the area of taxes, Finland, Sweden, Iceland and particularly Norway have traditions of disclosing individuals' income, wealth and tax payments to the public, with the implicit intention of increasing tax compliance (Bø et al., 2015).

In all cases described so far, it is important to emphasize that the information made public is not information about actual fraud, but only information about reports, their impact, or both. The effectiveness of these measures in reducing dishonest reporting is hard to quantify, as is often the case with topics concerning honesty (Bø et al., 2015). In schemes which combine the two aspects of lifting anonymity and making externalities clearly transparent (like in P2P-insurance), it is also hard to differentiate the impact of each of these two mechanisms on honesty.

We address this problem through a novel experimental task, the *claim game*. We present subjects in our lab experiments with envelopes containing either 30 or 70 euro-cents and allow them to claim the difference to 1 euro through an unverified self-report. Subjects, therefore, have the possibility to either claim the true difference to 1 euro or to submit a claim greater (or smaller) than the difference. To investigate the assumed effects described above, we expand upon this control treatment by introducing (1) a reduction in the anonymity of reporters by making unverified claims public, (2) an externality by paying claims from a pool jointly owned by a group of subjects, and (3) a combination thereof. This allows us to answer the question whether reducing anonymity about reporters and focusing on externalities of misreports affect honesty.

Our results show that removing anonymity significantly increases honesty (by up to 20 percentage points of average overclaiming), while adding the described type of externality does not. We find support for the hypothesis that the (no-)anonymity-effect is driven by activation of the wish to appear honest. An alternative explanation, intensified fear of experiencing shame, is not supported. Regarding the role of externalities, we find evidence that the fear of being harmed by excessive claims by other pool members decreases honesty and crowds out pro-social motives for being honest.

[1]Examples include Friendsurance in Germany and Guevara in the UK.

In the remainder of the article, section 2 discusses the relevant literature, section 3 details our design and methods, section 4 presents our results, and section 5 concludes with a discussion.

2. Literature review

The ample literature on honesty when submitting unverified self-reports shows that a significant number of participants are honest even in the absence of audits or danger of financial punishment (e.g., Fischbacher and Föllmi-Heusi, 2013). Abeler et al. (2016) propose a theoretical model for these results and test it in a meta-analysis of 72 experiments on the topic. From a range of hypotheses assumed to be potential explanations of the phenomenon, all but a combination of two are dismissed by the experimental data. The remaining explanations are the simultaneous existence of (1) a preference for being honest (pure lying aversion) and (2) a preference for being perceived as being honest. Abeler et al. (2016) define the first as causing an intrinsic cost when deviating from the truth (as stated, e.g., in Gneezy et al., 2013 and Fischbacher and Föllmi-Heusi, 2013), and the latter as causing an intrinsic cost depending on the perceived likelihood of appearing dishonest to an outside observer (a finding similar to, e.g., Hao and Houser, 2017, Hilbig and Hessler, 2013 and Fischbacher and Föllmi-Heusi, 2013).

Concerning variations in the observers of unverified self-reports, research so far has only varied whether the experimenter can see the true state or not (Abeler et al., 2016; Gneezy et al., 2016). Variation in the observability of only the report itself have to our knowledge not been studied. We know of three papers which vary anonymity in some way: Fischbacher and Föllmi-Heusi (2013) and Mazar et al. (2008) compare a standard payment procedure to a double-blind procedure in an honesty experiment, finding no differences in results. Conrads and Lotz (2015) vary anonymity by changing the communication channel between participant and experimenter from distant to close (for example from mail to face-to-face communication). They find that reducing distance increases reporting honesty in their setting.

Self-conscious emotions, particularly shame or guilt, are known to be guiding factors of moral behavior. The anticipation or fear of experiencing such emotions can lead to a higher degree of honesty. Greenberg et al. (2015) define guilt as the "desire not to let someone down in terms of monetary payoff" (just like Battigalli et al., 2013) and shame as a "desire to be perceived favorably". In a sender-receiver game experiment (after Gneezy, 2005), they find a strong effect of anticipated shame on honesty when informing senders whether the truthfulness of the reports they are going to send will afterwards be unveiled to receivers. Anticipated guilt plays a minor role in their setting. Coricelli et al. (2014) and Casal and Mittone (2016) find similar experimental results of increasing tax compliance when threatening to "shame" discovered tax evaders publicly. The degree of participants' shame proneness explains some of the observed behavior, while guilt, again, does not.

When using the broader psychological definition of self-conscious emotions, which defines guilt as a "negative evaluation of specific behavior" and shame as a "negative evaluation of the global self" (Tangney et al., 2007), both shame and guilt can operate without an audience and without disclosure of the actual offense. Rather than by how these emotions are triggered, they are differentiated by how they are experienced and what subsequent reactions they may trigger (remorse, sense of being worthless, etc.). From these two emotions, shame is usually experienced more strongly and negatively (Tangney et al., 2007), which would be in line with observed effects in the mentioned experiments. Particularly anticipated shame[2] and participants' proneness to this emotion could thus also play a role for honesty in games of unverified reports. This has yet to be studied.

Forms of social preferences should play no role in experiments of unverified self-reports. Gibson et al. (2013), for example, find that altruistic concerns are not relevant if there are no consequences for hypothetical victims of misreporting. This may be the reason why there is no research into the role of pro-social preferences in such situations. However, there are studies investigating the link between honesty and Social Value Orientation (SVO), a simple other-regarding preference capturing subjects' concern for the payoff of others. SVO is positively correlated with honesty both when reporting information to a hypothetical interaction partner (Rasmußen and Leopold-Wildburger, 2014) and when reporting the privately observed outcomes of die throws to the experimenter (Grosch and Rau, 2017).

In Gneezy's sender-receiver games, where a lie is targeted at another person, other-regarding preferences positively affect honesty (that is, the sender is more honest because she cares about her lie's effect on the receiver – cp. Gneezy, 2005; Hurkens and Kartik, 2009). Fischbacher and Föllmi-Heusi (2013) report on a dictator-like die experiment, where an amount of money is distributed between two players according to the self-report of one player's die roll. Reports (and thus, lies) decrease in the dictator setting, but not significantly.[3]

However, these situations differ from our externality setting in several respects: There are no substantial externalities in standard honesty experiments, since the only one hurt is the experimenter, who is perceived less as an actor in the experiment than as a part of the institution. While Gneezy's sender-receiver game includes externalities, they are not reciprocal. The receiver is always a purely passive victim and the sender the active perpetrator. In our reciprocal externality setting, a reporter affects others through her dishonest reports, but can also be affected herself by others' dishonest reports. The literature on corruption provides evidence that reciprocal externalities can in fact crowd out other-regarding preferences. Abbink et al. (2002) for example find that adding a reciprocal externality to a setting allowing for the possibility of bribery reduces neither the number of bribe offers nor the frequency of bribe acceptance. When "one-way" externalities are directed at a passive third party on the other hand, Barr and Serra (2009), in a similar setup, show that both bribe offer and acceptance rates decrease in the size of the externality. They hypothesize that this difference may be due to a decision-maker's belief that, under reciprocal externalities, others' behavior may hurt the decision-maker herself. In related work in the public goods

[2] In the sense of Tangney et al. (2007) shame and guilt are likely different from both, lying aversion and the preference for appearing honest.

[3] However, Fischbacher and Föllmi-Heusi (2013) themselves note the low statistical power of this specific experiment.

literature, Fischbacher and Gächter (2010) show that for participants strongly concerned about losing out, the fear of others harming a participant can outweigh the participant's pro-social motives. The fear of being exploited by others is furthermore correlated with dishonesty (Steinel and De Dreu, 2004).

Public good games are also instructive about behavior in settings where anonymity and reciprocal externalities are combined. This combination has a very strong effect on contribution levels, proving that non-monetary punishments like disapproval can have a similar effect as monetary fines (Masclet et al., 2003; Rege and Telle, 2004). Still, this situation does not accurately reflect our questions about anonymity, because transparency regarding reports in public goods games gives unambiguous information about the (mis)behavior of the reporter. Publicity of unverified reports in an honesty experiment always leaves room for doubt regarding the accuracy of the report and could, therefore, be more closely related to effects found in the bribery literature mentioned earlier.

While individual aspects of our research resemble elements of the experimental honesty, tax evasion, corruption and public goods literatures, these too do not offer unambiguous predictions about which effects to expect for our research. Our main contribution to the literature therefore is twofold: (1) We are the first to explicitly differentiate between lifting anonymity of reporters and between providing information about (reciprocal) externalities of unverified reports. We are not aware of research into behavioral drivers in such situations. As laid out, practitioners often seem to take as given a positive effect of reduced anonymity on honesty. (2) We add to the experimental honesty literature on unverified reports by investigating whether shame proneness, guilt proneness, and pro-social preferences drive honesty, specifically in the situations described earlier. Additionally, we also present a new experimental task, the *claim game*, which allows for studying honesty when group resources are affected by group members' reporting decisions.

Note that we also focus on an unframed setting and do not include audits, both of which – framing and audits – are often present in other experimental research on, e.g., tax evasion (Alm, 2012) or financial reporting fraud (Gibson et al., 2013). We believe this helps us exclude potential confounding factors and concentrate on the pure effects of anonymity and externalities on honesty.

3. Material and methods

3.1. Participants

We started by conducting a pilot experiment with 144 participants (138 after cleaning for errors) at the Max-Jung-Lab at the University of Graz. F-test power analysis of the resulting effect size of $f^2 = 0.031$ showed that a participant size of 79 participants per treatment was required to identify main effects and interactions, for a desired $\alpha = 0.05$ with power 0.8. Consequently, we conducted additional sessions in the Max-Jung-Lab at the University of Graz and in the EconLab at the University of Innsbruck (together with the pilot resulting in a total of 176 and 142 participants, respectively, after cleaning for errors). The final sample consisted of students with a mean age of 24.6 (SD=4.4), of which 55.3% were female. Participants were recruited via ORSEE in Graz and via hroot in Innsbruck

(Greiner, 2015; Bock et al., 2012). The experiment was computerized using z-Tree 3.6.7 (Fischbacher, 2007).

Given the brevity of the task (less than 20 minutes including SVO slider and other questionnaires), the sessions were always run as addenda to other experiments, but clearly indicated as being a separate task from the preceding experiments. We took portrait photos of participants at the beginning of each session. This part of our procedures was advertised in the invitation email for the experiment. Upon arrival in the lab, participants were explicitly asked whether they agreed to this, with the option of quitting the experiment immediately for a show-up fee of 3 euros (4 euros in Innsbruck, as per lab rules). All decided to participate.

On average, participants earned a total of 5.50 euros for participating in the experiment, in the SVO slider measure and for filling out the questionnaire. Data cleansing excluded a total of 10 (3.0% of total) participants for either observed or self-indicated errors. The exclusion of these participants does not significantly change our results (see Appendix E for details and analysis).

3.2. Design

We deploy a 2×2 between-subjects design to explore our research questions. We study two main effects: Anonymity, which refers to whether individuals' reports and pictures are visible to other participants, and externality, which relates to whether an individual's claim affects the payoffs of other participants in the experiment by draining a common pool. We see the latter as providing information about the impact of subjects' reports. Including the interaction between these manipulations, our treatments therefore are *Control* (anonymity, no externality), *Public* (no anonymity, no externality), *Ext* (anonymity, externality) and *PublicExt* (no anonymity, externality). All treatments have in common that we take a portrait photo of each participant upon entry to the Lab, independent of whether this photo is later used in the specific session or not. On their desks, participants find a sealed envelope containing either 30 or 70 euro-cents. We choose two levels of contents to investigate the impact of relatively low and relatively high contents (compared to the communicated maximum content of 100 cents) on honesty.

At the beginning of the experiment, participants are informed that every participant faces the same task and has either $0, 10, \ldots, 90$ or 100 cents in her envelope. Each participant is allowed to pocket the envelope content and does not need to show it to anyone else. We do not address whether anyone (specifically, the experimenter) knows the contents of the envelopes, but inform subjects that not everyone has the same envelope content. Since we are fully informed about envelope contents, we can identify behavior perfectly (see also the discussion of this design choice at the end of this subsection). Participants are also informed that everyone is entitled to receive a total of 1 euro from this part of the experiment, which they can obtain by claiming the difference between their envelope content and 1 euro in the main stage of the experiment. There, participants are asked to enter their claim $\in \{0, 10, ..., 90, 100\}$, which is paid out to them at the end of the experiment. In the following, we will refer to this part of the experiment as the *claim game*.

Following the claim game, we elicit social preferences and proneness to guilt and shame, which the participants are compensated for separately from the payoff of the claim game.

Specifically, we measure the social preferences of our subjects using the SVO Slider measure (Murphy et al., 2011; Murphy and Ackermann, 2013). The six primary items are dictator game-like distribution decisions between oneself and an anonymous other with varying marginal rates of substitution. This yields an SVO-angle $\in [-16.26^o, 61.39^o]$ indicating a subject's concern for the other person's payoff. By taking its tangent, the angle can be transformed into the parameter $\alpha \in [-0.29, 1.83]$ of an other-regarding social preference function $U(\pi_s, \pi_o) = \pi_s + \alpha \cdot \pi_o$. Here, π_s is the payoff for the self and π_o is the payoff for the other. A greater SVO-angle indicates more concern for others' payoffs and thus is an indicator for more pro-social preferences. We use the z-Tree implementation of the SVO slider in the German version by Crosetto et al. (2012), with a self-written introduction as laid out in Appendix A. We furthermore use the TOSCA-3 test of self-conscious affect (Tangney et al., 2000) to measure proneness to shame, guilt, externalization, and unconcern, of which we are interested only in the commonly used scales of shame and guilt (both $\in [11, 55]$). The test briefly describes different everyday situations and possible reactions thereto. Subjects then state the probability with which they would show, in this situation, each of the reactions described. Higher scores are related to higher proneness to the relevant emotion. We use the German TOSCA-3 short version consisting of 11 scenarios (also used by Rüsch et al., 2007).

The treatment *Control* follows the experiment exactly as described. For the other treatments, we apply the following changes: We inform participants at the beginning that they will be randomly assigned to groups of 4. Participants in treatment *Public* are additionally informed that after filing the claim, all participants in their group will see the photos of all group members together with their claims (but not their envelope content) before receiving the payoff (inspired by Coricelli et al., 2014). Figure 1 provides an example of this screen.[4] Subjects are also informed that this stage does not influence their payoff. Participants in treatment *Ext* are informed that their claims will be taken out of a common pool, jointly owned by the four (anonymous) members of their group. The pool contains 4 euros. This is enough to satisfy all claims, even if everyone filed the highest possible claim. The money left in the pool after all claims have been satisfied is distributed in equal shares to the owners of the respective pool. Obviously, by overclaiming individually, the overclaimer reduces the payoff of the other group members.[5] There is no information regarding the identity of the group members, so no offender or victim can be suspected at an individual level. In the final treatment, *PublicExt*, participants are subject to both the no-anonymity and the externality conditions as described earlier. They are informed that they will make their claims from the common group pool and that the claim of each group member will be displayed, together with their pictures, to all group members after everyone has made their claims. The sequence of the experiment is summarized in Figure 2. Detailed instructions to participants for each stage can be found in Appendix A.

With this design, it is possible to separately study the effects of lifting reporters' anonymity

[4]Note that the example photos do not picture actual participants, but the authors and an author's spouse.

[5]The marginal per-capita return (MPCR) of leaving money in the pool is $1/n = 0.25$. Rational, exclusively self-interested subjects thus have no incentive to claim less than 1 euro.

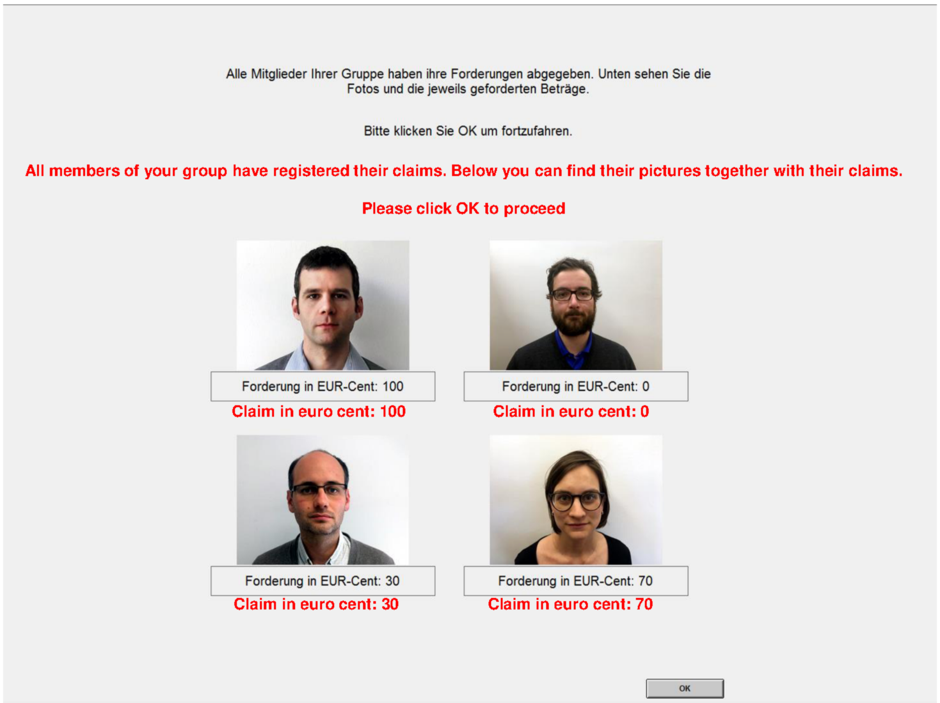

Figure 1: Screen used in transparency stage of experiment. English translations added in red font (grey in grey-scale printouts) were not part of the original screen.

and of creating transparency about externalities, as well as of the interaction of these two manipulations. Furthermore, since subjects claim the difference relative to a fixed amount, fairness considerations about unequal initial endowments should not be a reason for dishonest reporting. If everyone in the experiment were honest, everyone would receive exactly the same amount (that is, a total of 1 euro + equal payouts from the pool, where applicable). We see this as an advantage over classical honesty experiments, where we suspect that subjects have the bad luck to observe a low payoff state in private may be more likely to lie out of a preference for fairness. They may perceive their lie to merely correct an unfair initial condition relative to that of the other participants, particularly in a setting with reciprocal externalities.

Note that we explicitly design the experiment to make individuals' decisions about honesty observable (for the experimenter). Most researchers on the topic of honesty in unverified reports rely on the "die paradigm" (Fischbacher and Föllmi-Heusi, 2013). This approach guarantees anonymity to reporters but allows researchers to judge aggregate dishonesty only. We believe that our design is necessary to clearly identify effects in the situations we target. The literature contains mixed results on whether observability of the true state substantially impacts results in honesty experiments. Gneezy et al. (2016) for example run a reporting experiment varying whether the truthfulness of reports is observable. They find a significant treatment difference in the number of dishonest reports, but a substantial number of dishonest reports still persist when outcomes are made observable to the experimenter. The most obvious change between treatments is a decline in the number of "partial" liars

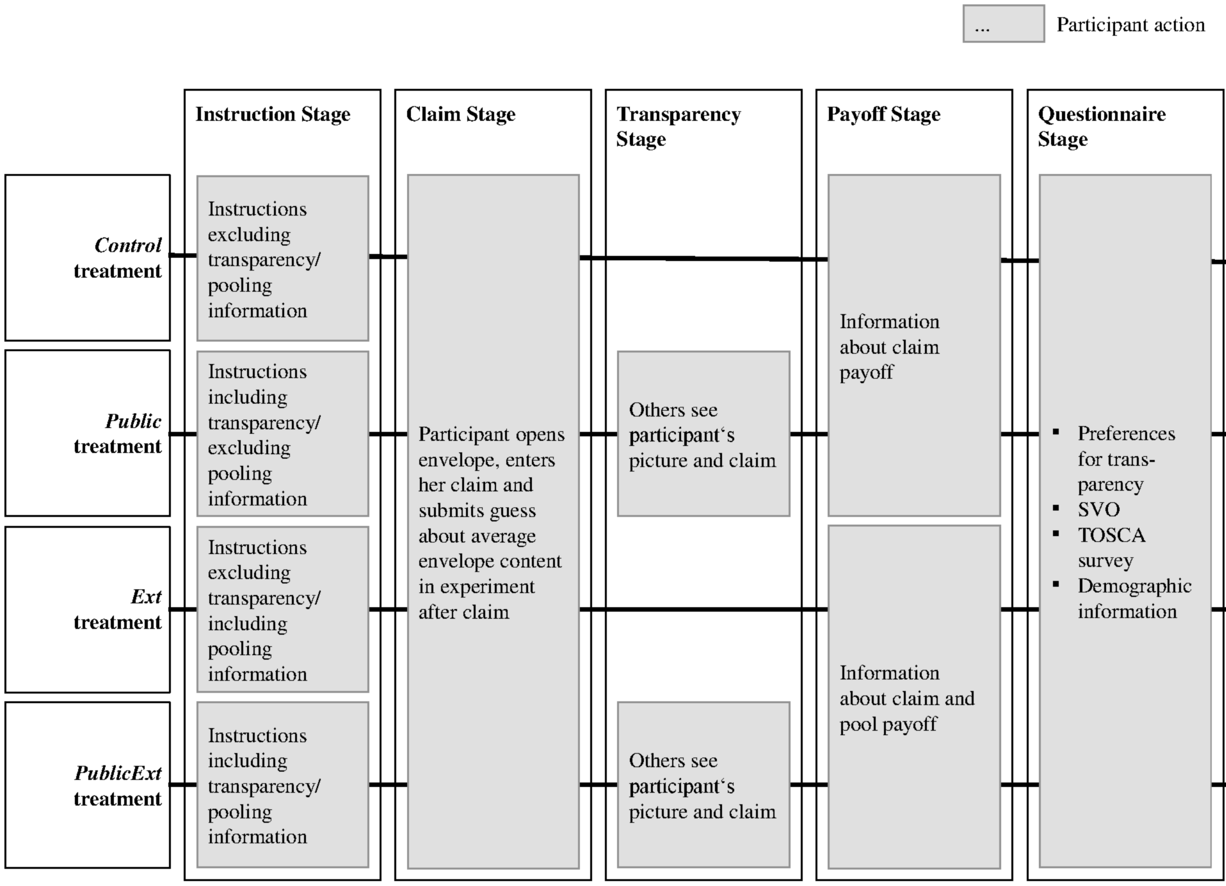

Figure 2: Experiment stages by treatment.

(those who lie but do not claim the maximum) under observability. Abeler et al. (2016) find similar results, even with slightly stronger effects. Quite to the contrary, van de Ven and Villeval (2015) report no significant differences in honesty when adding an observer to a sender-receiver game. Note, though, that in sender-receiver games all actions are always transparent for the experimenter anyway. We believe that the behavioral changes identified under observability in Gneezy et al. (2016) and Abeler et al. (2016) hail mainly from two sources: (1) The more effort is put into providing anonymity, and, therefore, the more attention is drawn to questions about anonymity, the more likely is an increase in honesty, and (2) the easier it is to connect the setup with observability (e.g., using a computerized random device whose outcome is clearly recorded), the more likely is an increase in honesty. We thus do not make any reference to observability or anonymity of payments and use a physical device (money in an envelope) with a certain psychological separation from the PC in the hope for valid results despite the identifiability of dishonest reports. We find that our results bear out this hope. As reported in section 4, we find substantial proportions of partial and full overclaimers (21.3% and 31.3%, respectively) in the *Control* treatment, which in total do not differ much from the estimated 33% partially (upper bound) and 18.3% fully dishonest people from the original baseline die experiment of Fischbacher and Föllmi-Heusi (2013).[6] All of our other treatments also exhibit a fair number of both full and partial overclaimers.

3.3. Hypotheses

We are foremost interested in the results of our main effects from the anonymity and externality treatments, as well as from the interaction between the two. We thus consider them in hypotheses H1 to H3. We then focus on the main effects of shame proneness, guilt proneness and social preferences (SVO) on honesty in H4 through H6. Finally, we present hypotheses on interactions between treatments and preferences in H7 through H9.

In the following, when speaking of differences in the extent of dishonesty, we mean differences in average claim amounts in euros. Similarly, when speaking of differences in the numbers of cases of dishonesty, we mean differences in the number of participants claiming more than they are entitled to.

As increased scrutiny by additional observers likely either increases anticipated shame, the wish to appear honest, or both, we postulate:

H1 Lifting anonymity decreases both the number of cases and the extent of overclaiming.

The evidence from the literature regarding the effect of externalities is less clear. In deriving our hypotheses, we rely on the fact that social preferences are generally seen as a driving factor of honesty when interacting with others (Gneezy, 2005; Grosch and Rau, 2017). However, under reciprocal externalities (as is the case here), other authors find that harmful behavior is not reduced (Abbink et al., 2002). Finally, in public goods games, the fear of losing out when others do not contribute can have detrimental effects on cooperation (Fischbacher and Gächter, 2010). Depending on their relative strengths, these two effects may cancel out or one of the following two competing hypotheses may hold:

H2a Adding externalities decreases both the number of cases and the extent of overclaiming.

H2b Adding externalities increases both the number of cases and the extent of overclaiming.

[6] We estimate these values from Fischbacher and Föllmi-Heusi (2013) as follows: Participants rolled a die in private, implying 6 outcome states (1,2,3,4,5,6) with respective payoffs (1,2,3,4,5,0) in CHF. Assuming no one reports to one's disadvantage, we take the fraction of subjects who reported the lowest payoff yielding 0 CHF (6.4%) as an estimate for the lower bound of the fraction of honest participants in all but the maximum payoff state. In the maximum payoff state, yielding 5 CHF, the expected fraction (16.7%) is taken as an estimate for the fraction of honest participants (as we assume that no one will lie downwards). In total, this yields 48.7% honest participants. The difference between the observed and the expected fraction in the maximum payoff state is estimated to be the percentage of fully dishonest participants (18.3%). The remainder (33%) is classified as partially dishonest. In our view, this is likely an upper bound for partial liars, as honesty in higher payoff outcome states could increase when the marginal profit from deviating becomes insufficient to compensate for the marginal increase in the psychological cost of appearing increasingly more dishonest.

We expect an even stronger effect on honesty when externalities are combined with a public display of reports. In this case, all aspects of increased scrutiny come into play on top of the effects of pro-social preferences alone. On the other hand, again, we cannot predict how strong the fear of losing out due to others' actions may be. As a benchmark, we thus posit the following hypothesis, which assumes that the effects from the two main effects are purely additive:

H3 There are no interactions between the effects of lifting anonymity and of adding externalities on either the number of cases or the extent of overclaiming.

While not our main focus, we also propose hypotheses about the main effects of moral emotions and pro-social preferences. Coricelli et al. (2014) and Casal and Mittone (2016) use the same measure as we do to assess proneness to feeling shame and guilt. We rely on their experimental findings to inform our expectations:

H4 The number of cases and the extent of overclaiming decrease in subjects' proneness to feeling shame.

H5 The number of cases and the extent of overclaiming do not vary in subjects' proneness to feeling guilt.

As mentioned in section 2, pro-social subjects (as measured using the SVO slider measure) are more honest both when reporting information to a hypothetical other person and in the die task (Rasmußen and Leopold-Wildburger, 2014; Grosch and Rau, 2017). Therefore, and in line with research relating SVO to various pro-social behaviors like pro-environmental choices (Joireman et al., 2001) and donations to charity (Van Lange et al., 2007), we conjecture:

H6 The number of cases and the extent of overclaiming decreases in the strength of subjects' pro-social preferences.

The literature does not provide a clear indication about what may be the main drivers of interaction effects between treatments and preferences. We argue that the presence of observers may intensify the experience of shame and may increase the fear of experiencing this emotion. Therefore we expect a stronger effect of shame proneness under our treaments with no anonymity:

H7 The effect of subjects' proneness to feeling shame on the number of cases and the extent of overclaiming is stronger in *Public* and *PublicExt*.

We do not make any further assumptions regarding the proneness to feeling guilt, as we do not expect it to have an effect on honesty (as posited in hypothesis H5). Pro-social preferences, on the other hand, should play a more important role when subjects face actual victims who are hurt by their reports, since pro-social subjects derive utility from others' payoffs. Therefore:

H8 The effect of subjects' pro-social preferences on the number of cases and the extent of overclaiming is stronger in *Ext* and *PublicExt*.

We do not have a way to directly measure lying aversion and the preference for appearing honest (assuming they are different from anticipated guilt and shame). Still, recently proposed theoretical models, such as those in Abeler et al. (2016), Dufwenberg and Dufwenberg (2016), Gneezy et al. (2016) and Khalmetski and Sliwka (2017) indicate that the effect on honesty of lifting anonymity is driven by a preference for appearing honest. Following this line of argument, an increase in image concerns and, thereby, the wish to appear honest should be reflected in specific changes of behavior in our setting as follows:

Imagine a participant in our experiment. With anonymity, she will weigh her profit from claiming a certain amount against her intrinsic lying costs (which increase with the deviation from her true claim) and against her cost from the perceived probability of appearing dishonest (which increases as her claim approaches the maximum claim). Under the assumption that lifting anonymity does not affect the participant's intrinsic lying costs, but only affects the desire to appear honest, this added pressure will decrease her report. However, participants will refrain from underclaiming, such that the amount they are truly entitled to acts as a floor for their possible claim amounts.[7]

This has the following consequence: Ruling out underclaims, a participant with an envelope content of 30, who should honestly claim 70 cents, can overclaim by between 0 and 30 cents only. A participant with an envelope content of 70, who should honestly claim 30 cents, can overclaim by between 0 and 70 cents. Now assume that observers judge the appearance of honesty by comparing the claim amount to, e.g., the middle of the range of possible claim amounts, i.e., 50 cents. Compared to this reference point, participants with an envelope content of 30 cents are likely to appear relatively dishonest to observers even when claiming only the amount of 70 cents which they are entitled to. Such participants would have to accept a steep price in terms of the psychological costs of appearing dishonest were they to overclaim. A participant with an envelope content of 70 cents, conversely, could easily overclaim by a few 10 cent-steps before appearing overly suspect of dishonesty. Assuming a preference for appearing honest, this implies a greater frequency of honest claims under an envelope content of 30 than of 70 cents. It also implies an increase in partial overclaimers under an envelope content of 70 compared to a content of 30 cents, at the expense of full overclaimers in the no-anonymity treatments.

Turning this into a hypothesis, we therefore postulate:

H9 The number of cases and the extent of (both partial and full) overclaims under no anonymity decreases more for envelope content 30 than 70 cents, compared to the situation under anonymity. This is driven by a reduction in full overclaimers which

[7] The assumption relies on a disproportionally high cost of claiming less than what a participant is entitled to. A participant who underclaims would both lie and lose out on money she is entitled to. Thus, her report is bounded from below at the honest claim, which is the difference between 1 euro and her envelope content. This prediction of "no underclaimers" is borne out by observations in games with full observability of reports such as in Gneezy et al. (2016) and Abeler et al. (2016).

is comparable under both envelope contents, but an increase in partial overclaimers which is greater under envelope content 70 than 30 cents.

Note that we make no assumptions about the direct effect of envelope content on dishonesty. While we believe to be able to gain insights from the interaction between envelope content and anonymity, we do not think a variation of the influence of envelope content on its own to be instructive. Therefore, we will only interpret interactions between envelope content and the main effect of anonymity.

4. Results

4.1. Treatment effects

We start by reporting descriptive statistics on claiming behavior. Let PARTICIPANT_CLAIM be the claim registered by a participant and ENVELOPE_CONTENT the amount she finds in her envelope. Then, DEVIATION is the deviation (in euro-cents) of the participant's actual payoff from the 100 cents she is entitled to:

$$\text{DEVIATION=PARTICIPANT\_CLAIM+ENVELOPE\_CONTENT-100}$$

Given ENVELOPE_CONTENT $\in \{30, 70\}$, DEVIATION lies in $\{-70, -60, \ldots, 30\}$ for the first and $\{-30, -20, \ldots, 70\}$ for the second content, respectively. It follows that DEVIATION $\in \{-70, -60, \ldots, 70\}$. We now classify participants into 5 behavioral types:

$$\text{CLAIM\_TYPE} = \begin{cases} \text{Full OC,} & \text{if PARTICIPANT\_CLAIM} = 100 \\ \text{Partial OC,} & \text{if PARTICIPANT\_CLAIM} < 100 \text{ and DEVIATION} > 0 \\ \text{Honest,} & \text{if DEVIATION} = 0 \\ \text{Partial UC,} & \text{if PARTICIPANT\_CLAIM} > 0 \text{ and DEVIATION} < 0 \\ \text{Full UC,} & \text{if PARTICIPANT\_CLAIM} = 0 \end{cases}$$

CLAIM_TYPE thus defines the direction and extent of underclaiming (UC) or overclaiming (OC), meaning whether the deviation is below or above the entitlement of 1 euro.

Table 1 shows the final number of participants in each treatment, split by CLAIM_TYPE. Overall, we observe substantial fractions of both partial (21.4%) and full overclaimers (21.4%) across all treatments. Surprisingly, we also observe a few partial and even full underclaimers (6.3% and 0.9%, respectively) which we did not expect and therefore did not include in our hypotheses. As this is an unusual observation compared to the literature, we discuss these reports separately in subsection 4.4.

We analyze the main effects using the binary variables PUBLIC (true in treatments *Public* and *PublicExt*, false otherwise) and EXTERNALITY (true in treatments *Ext* and *PublicExt*, false otherwise). We find substantial differences in the number of full overclaimers when splitting by PUBLIC. Figure 3 shows that lifting anonymity more than halves the number of full overclaimers (proportions test $\chi^2 = 10.27$, $p = 0.001$). PUBLIC and EXTERNALITY

Treatment	*Control*		*Public*		*Ext*		*PublicExt*		Total (Row)	
	#	(%)	#	(%)	#	(%)	#	(%)	#	(%)
Full OC	25	(31.3)	9	(10.8)	21	(26.9)	13	(16.9)	68	(21.4)
Partial OC	17	(21.3)	22	(26.5)	11	(14.1)	18	(23.4)	68	(21.4)
Honest	34	(42.5)	44	(53.0)	40	(51.3)	41	(53.3)	159	(50.0)
Partial UC	4	(5.0)	8	(9.6)	4	(5.1)	4	(5.2)	20	(6.3)
Full UC	0	(0.0)	0	(0.0)	2	(2.6)	1	(1.3)	3	(0.9)
Total (Col)	80		83		78		77		318	

Table 1: Descriptive statistics of number of each CLAIM_TYPE by treatment. Percentages are fractions of column totals.

each have a positive but limited effect on the total number of full and partial overclaimers combined. These effects are, however, not significant (proportions tests with $p > 0.1$ for both).

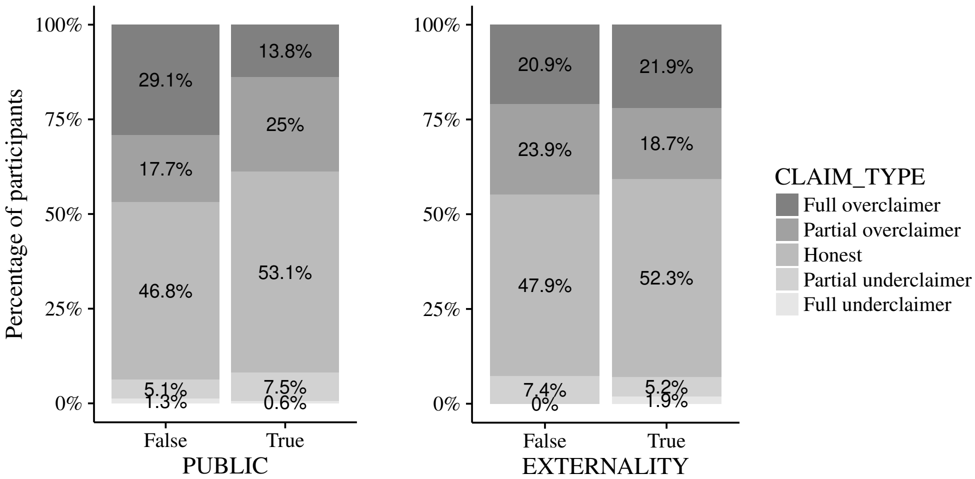

Figure 3: Percentages of claim types, split by either main effect PUBLIC or EXTERNALITY.

To easily assess the extent of dishonesty in claims across envelope contents, we also calculate the following measure:

$$\text{RELATIVE\_DEVIATION} = \begin{cases} 100 \cdot \frac{\text{DEVIATION}}{100 - \text{ENVELOPE\_CONTENT}}, & \text{if DEVIATION} < 0 \\ 0, & \text{if DEVIATION} = 0 \\ 100 \cdot \frac{\text{DEVIATION}}{\text{ENVELOPE\_CONTENT}}, & \text{if DEVIATION} > 0 \end{cases}$$

RELATIVE_DEVIATION is DEVIATION measured relative to the maximum possible deviation in the direction of the deviation. To make this number easier to interpret, we use a scaling factor of 100, such that the measure can be interpreted as the percentage deviation in either direction relative to the maximum possible deviation determined by

ENVELOPE_CONTENT. As a result, RELATIVE_DEVIATION $\in [-100, 100]$.[8] The rationale for using RELATIVE_DEVIATION instead of DEVIATION is to facilitate comparison without the need for a split by envelope content. We will only report results on RELATIVE_DEVIATION from here on, yet we provide a comparison of the two measures, and we provide the key econometric model for both DEVIATION and RELATIVE_DEVIATION in Appendix C.

Figure 4 shows the interaction graph of average RELATIVE_DEVIATION. While lifting anonymity significantly reduces the amount of overclaiming without externalities, the situation is unclear after adding externalities. There is some suggestion of an interaction that reduces the effect of lifting anonymity when coupled with a reciprocal externality.

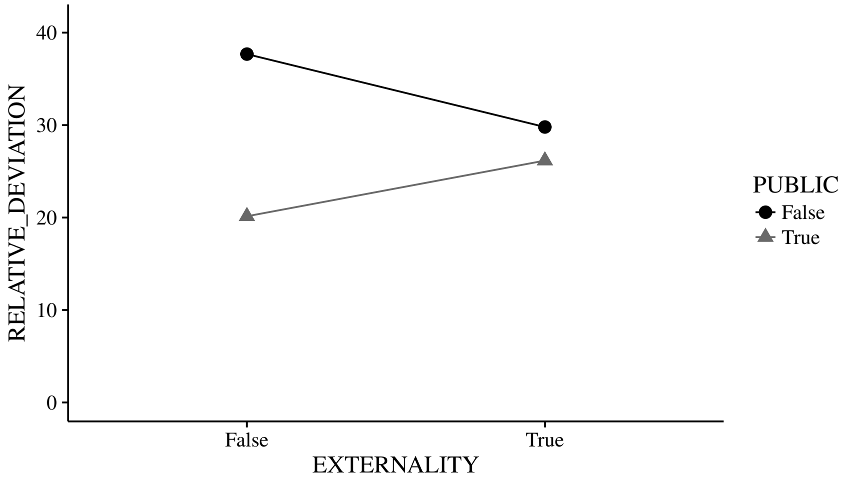

Figure 4: Interaction plot of RELATIVE_DEVIATION by main effects PUBLIC and EXTERNALITY.

For a combined analysis of our hypotheses on the number of cases of overclaiming (e.g., the likelihood of being of a certain claim type), we use an ordinal regression to study CLAIM_TYPE (coded 0 for Full UC, 1 for Partial UC, 2 for Honest, 3 for Partial OC and 4 for Full OC). As independent variables we include our main effects PUBLIC and EXTERNALITY (each coded 0 for False and 1 for True), their interaction, SHAME_SCORE, GUILT_SCORE, SVO_ANGLE, and ENVELOPE_CONTENT_70 (coded as 0 for 30 cents and 1 for 70 cents). To investigate changes in the extent of lying, we furthermore run a Tobit regression with RELATIVE_DEVIATION as the dependent variable, left- and right-censored at -100 and 100, respectively. We use the same independent variables as for the ordinal regression.

Table 2 shows the results of the regressions in the columns without interactions. They picture is very consistent across models. We find a significantly positive effect of lifting anonymity, which decreases RELATIVE_DEVIATION by more than 20 percentage points.

[8] Strictly speaking, RELATIVE_DEVIATION can also only take on discrete values in the interval, but we abstract from this to simplify notation.

	Ordinal		Tobit	
	No interactions	Interactions	No interactions	Interactions
INTERCEPT			116.38^{***}	52.18^{*}
			(24.98)	(27.67)
PUBLIC	-0.84^{***}	-0.91^{**}	-22.43^{***}	-23.63^{**}
	(0.30)	(0.38)	(8.46)	(10.31)
EXTERNALITY	-0.38	-0.36	-7.20	-7.02
	(0.32)	(0.32)	(8.69)	(8.70)
PUBLIC × EXTERNALITY	0.64	0.63	14.40	14.09
	(0.44)	(0.44)	(12.08)	(12.11)
SHAME_SCORE	-0.03^{**}	-0.03	-0.98^{**}	-0.86
	(0.02)	(0.02)	(0.43)	(0.58)
SHAME_SCORE × PUBLIC		-0.02		-0.27
		(0.03)		(0.79)
GUILT_SCORE	-0.01	-0.01	-0.33	-0.32
	(0.02)	(0.02)	(0.61)	(0.61)
SVO_ANGLE	-0.05^{***}	-0.06^{***}	-1.46^{***}	-1.50^{***}
	(0.01)	(0.01)	(0.22)	(0.31)
SVO_ANGLE × EXTERNALITY		0.00		0.11
		(0.02)		(0.43)
ENVELOPE_CONTENT_70	0.76^{***}	0.69^{**}	16.68^{***}	15.14^{*}
	(0.22)	(0.32)	(6.04)	(8.71)
ENVELOPE_CONTENT_70 × PUBLIC		0.15		3.06
		(0.44)		(12.19)
AIC	730.31	735.86	2823.12	2828.90
BIC	771.69	788.53	2856.98	2874.04
Log Likelihood	-354.16	-353.93	-1402.56	-1402.45
Deviance	708.31	707.86	395.79	395.89
Total	318	318	318	318
Left-censored			3	3
Uncensored			247	247
Right-censored			68	68
Wald Test			69.65	70.00
McFadden Pseudo R^2	0.090	0.091	0.023	0.023

$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

Table 2: Ordinal regression of CLAIM_TYPE and Tobit regression of RELATIVE_DEVIATION (left censored at -100 and right censored at 100), both with and without interactions. SHAME_SCORE and SVO_ANGLE are mean centered in models with interactions.

Result 1 *Lifting anonymity by making subjects' claims public significantly reduces the number of full overclaims (by half) and reduces RELATIVE_DEVIATION from the true claim by more than 20 percentage points. Hypothesis 1 is supported.*

Table 2 also shows that the coefficient of EXTERNALITY is insignificant (albeit negative, as expected). The interaction PUBLIC × EXTERNALITY points in the same direction as observed in the interaction graph in Figure 4, with externalities partly canceling out the effect of lifting anonymity. However, the coefficient is also insignificant. This supports H1 and H3, while failing to offer support for H2a and H2b, as we observe no significant effect in either direction from EXTERNALITY.

Result 2 *Reciprocal externality neither increases nor decreases the number of overclaimers and does not affect RELATIVE_DEVIATION. Hypotheses 2a and 2b are not supported. There is no significant interaction between reciprocal externalities and anonymity. Hypothesis 3 is supported.*

4.2. Explanatory factors for observed effects

As expected, stronger pro-social preferences and higher proneness to feeling shame are significant indicators for a lower degree of dishonesty. As Table 2 shows, they reduce the probability of becoming a partial or full overclaimer and they reduce RELATIVE_DEVIATION. The coefficient of GUILT_SCORE is far from being significant, even if it has the same sign as that of SHAME_SCORE. This supports H4 through H6.

Result 3 *Stronger pro-social preferences and higher proneness to feeling shame are correlated with greater honesty, while proneness to feeling guilt shows no effect. Hypotheses 4 through 6 are supported.*

For our next analysis, we add (mean centered) interactions between PUBLIC and SHAME_SCORE as well as between EXTERNALITY and SVO_ANGLE to the regression models to study hypotheses H7 and H8 about the effects of individual scores in the respective treatments. We also add an interaction between PUBLIC and ENVELOPE_CONTENT_70 to assess hypothesis H9. The results are reported in the models with interaction in Table 2. Contrary to our initial hypothesis H7, SHAME_SCORE does not interact significantly with PUBLIC. In fact, adding the interactions negatively affects the significance of the main effects of PUBLIC and SHAME_SCORE, but does not add predictive power (McFadden's Pseudo R^2 remains almost unchanged).

Result 4 *There is no evidence that proneness to feeling shame varies in anonymity. Hypothesis 7 is not supported.*

Table 2 also does not document a significant effect of EXTERNALITY on honesty, nor is the effect of pro-social preferences significantly stronger or weaker in the externality treatments.

Result 5 *There is no evidence that the presence of externalities strengthens the effect of prosocial preferences. Hypothesis 8 is not supported.*

		BEING_HARMED											
		1		*2*		*3*		*4*		*5*		*Total*	
HARMING_ OTHERS	*1*	25	(16%)	2	(1%)	1	(1%)	4	(3%)	3	(2%)	35	(23%)
	2	1	(1%)	7	(5%)	5	(3%)	4	(3%)	1	(1%)	18	(12%)
	3	0	(0%)	2	(1%)	16	(10%)	4	(3%)	6	(4%)	28	(18%)
	4	5	(3%)	10	(6%)	2	(1%)	5	(3%)	1	(1%)	23	(15%)
	5	4	(3%)	6	(4%)	8	(5%)	6	(4%)	27	(17%)	51	(33%)
	Total	35	(23%)	27	(17%)	32	(21%)	23	(15%)	38	(25%)	155	(100%)

Table 3: Distribution of answers to questions "Have you thought about the fact that false report from others could hurt you?" (BEING_HARMED) and "Have you thought about the fact that false reports of you could hurt others?" (HARMING_OTHERS). 5-point Likert scale with 1="Not at all" to 5="Fully". Cells below the main diagonal (green) are in the category Self<Other, cells on the diagonal (yellow) are in Self=Other and cells above the diagonal (red) are in Self>Other. Percentages in brackets are fractions of total participants in treatments *Ext* and *PublicExt* combined.

According to the bribery literature (e.g., Barr and Serra, 2009), one reason for this result could be that the externalities in the claim game are reciprocal: Participants worry about being harmed by others and about harming others with their own actions. If these concerns are strong, the situation could change from an interaction driven by honesty aspects to one more strongly driven by strategic considerations regarding the behavior of other group members. At the conclusion of the externality treatments (*Ext*, *PublicExt*) we therefore ask participants to indicate, on a 5-point scale (1="Not at all" to 5="Fully"), whether they thought about (1) how other participants' overclaiming could hurt them (BEING_HARMED), and (2) how they would hurt others by overclaiming (HARMING_OTHERS). There is no difference in the distributions of answers between *Ext* and *PublicExt* (Kolmogorov-Smirnov with $p > 0.1$ for both questions), which suggests that anonymity does not materially influence these considerations and we do not need to consider this aspect separately. Table 3 shows the distribution of the two scores, combined over both externality treatments. It documents slightly stronger extremes in the HARMING_OTHERS scale and focal points at the extremes (1,1) and (5,5), followed by the neutral position (3,3). More than 30% of the participants therefore belong to one of 2 types: those who consider externality effects intensively are in (5,5) and those who do not consider them at all are in (1,1). However, RELATIVE_DEVIATION does not differ significantly between these two types (average RELATIVE_DEVIATION is 28.4 for (5,5) and 42.3 for (1,1), Wilcoxon rank sum test $W = 295$, $p = 0.397$).

Thoughts about harming others are weakly correlated with SVO_ANGLE (Spearman's $r = 0.175$, $p = 0.030$), which is consistent with the definition that pro-social subjects care about others' payoffs in this setting. However, thoughts about being harmed are independent of SVO_ANGLE (Spearman's $r = -0.02$; $p = 0.805$).

Connecting these considerations about self and others, we interpret cases where thoughts about harming others outweigh thoughts about being harmed (in the sense that participants indicated stronger thoughts about the one than the other) to indicate that pro-social thoughts outweigh the fear of losing out. Given this interpretation, we split participants with EXTERNALITY=True into three groups: those who report having stronger thoughts about harming others than about being harmed, those with equally strong thoughts and those with weaker thoughts about harming others, as indicated by position and color in Table 3 (below the diagonal or green, on the diagonal or yellow, above the diagonal or red). We then analyze whether dominance of one type of thought over the other influences reports at the individual level. Figure 5 suggests that this is the case. Those thinking more strongly about being harmed than about harming are more likely to be overclaimers than others. However, differences between the groups are not significant (χ^2 test of independence $\chi^2 = 3.60$, $\alpha = 0.463$), which could be related to the small sample size in each cell (Post-hoc power for a χ^2-test at the observed effect size of 0.11 and a desired $\alpha = 0.05$ is only 0.14).

In total, we do not have entirely convincing explanations for the absence of an externality effect. We do provide indirect support for the conjecture that beliefs about others' behavior may be driving decisions in this treatment, and that this effect may be strong enough to crowd out some of the effect of PUBLIC.

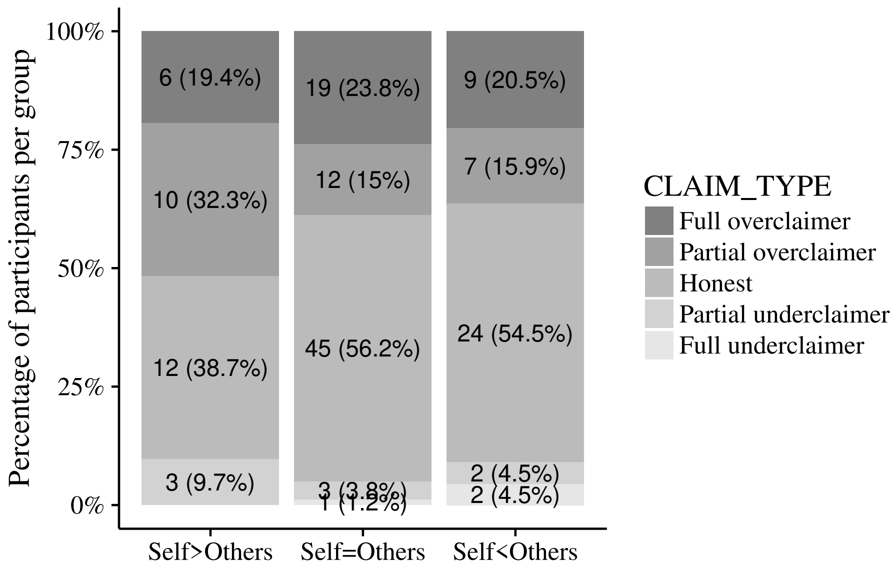

Figure 5: Fraction of CLAIM_TYPE by whether thoughts about being harmed were reported as being stronger than (Self>Others), equal to (Self=Others) or weaker than (Self<Others) thoughts about harming others. Restricted to treatments *Ext* and *PublicExt*.

Concerning ENVELOPE_CONTENT we are not interested in the main effect, but only in the interaction with PUBLIC (as laid out in the reasoning behind H9 in subsection 3.3). As expected, the interaction is positive, yet it is not significant (Tobit coefficient $p = 0.802$, ordinal coefficient $p = 0.586$; see Table 2).

Focusing directly on the distribution of CLAIM_TYPE by ENVELOPE_CONTENT in Figure 6, we predicted the significant increase in honest reports for ENVELOPE_CONTENT=30 when changing from PUBLIC=False to True in H9 (one-sided proportions test $\chi^2 = 3.07$, $p = 0.040$). For ENVELOPE_CONTENT=70 we see no significant increase in honesty (one-sided proportion test $\chi^2 = 0$, $p = 0.500$), but – also as predicted – a substantial change from full to partial overclaimers. Both results are in line with our reasoning behind H9. A Chochran-Mantel-Haenszel-Test for differences in the effects between envelope contents diagnoses significance at $\alpha = 0.1$ (one-sided $\chi^2 = 1.79$, $p = 0.090$). We attribute the lack of greater significance to low power after splitting the data into another two groups (Post-hoc power for a one-sided test of proportions at the observed effect size of 0.30 and a desired $\alpha = 0.05$ is 0.61 for ENVELOPE_CONTENT=30). Given the direct evidence for H9 and the problems of low power when introducing this many interactions, we conjecture that H9 may find greater support if one were to use a bigger sample.

Result 6 *There is a positive but insignificant effect of the interaction between anonymity and envelope content on overclaiming. We refrain from rendering conclusive judgment on Hypothesis 9 despite the lack of significance, since we also find a large amount of indirect supporting evidence. The evidence concerning Hypothesis 9 is mixed.*

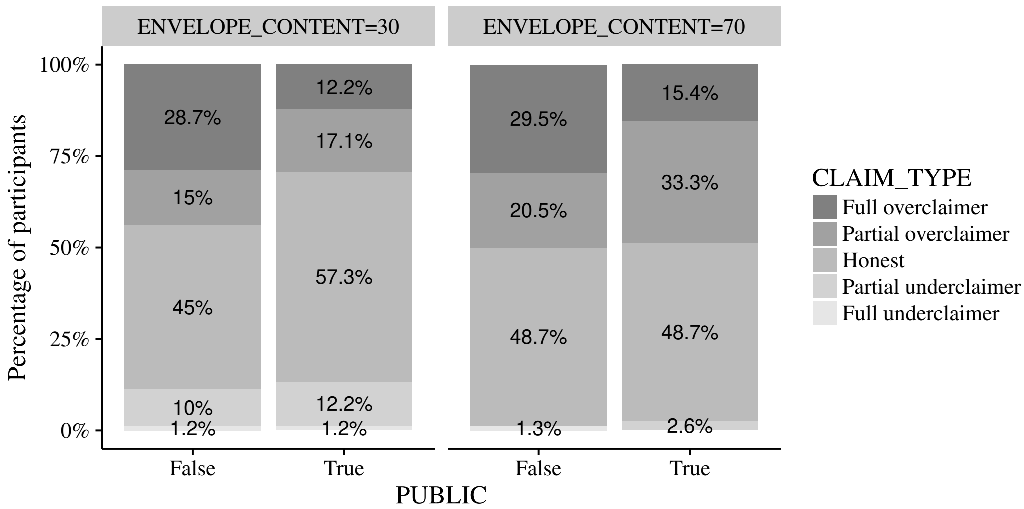

Figure 6: Percentages of CLAIM_TYPE, split by ENVELOPE_CONTENT and main effect ANONYMITY.

4.3. Robustness and controls

We explore the robustness of our findings in a number of ways. Adding a series of controls to the regression models causes no changes in any of the previously obtained results. Our results are furthermore robust to using other econometric methods (see Appendix B for details).

However, Table B.5 shows that four controls turn out to be (at least marginally) significant when we try to explain CLAIM_TYPE and RELATIVE_DEVIATION: a participant's estimate of average envelope content, the self-indicated preference for anonymity in the given reporting situation, the laboratory in which the experiment was run and the profit obtained in the experiments preceding the claim game. Concerning the marginal significance of the envelope estimates ($b = -0.47$, $p = 0.072$ for Tobit and $b = -0.02$, $p = 0.010$ for ordinal regression), note that participants report their beliefs about average envelope content in the claim game after opening their envelopes and entering their claims, but before receiving information regarding their payoffs. It seems plausible that participants who care for their appearance claim less if they believe average envelope content to be high. The reason is that, assuming the other participants are honest, high average envelope contents would be associated with low average claims. An overclaim would thus be more suspicious than when average envelope contents are low (and average claims high). We see this relationship with caution though, as there is essentially no correlation between ENVELOPE_GUESS and RELATIVE_DEVIATION (Spearman's $r = -0.034$, $p = 0.548$) and we also do not observe different correlations for the cases where PUBLIC=True (Spearman's $r = -0.036$, $p = 0.648$) and where PUBLIC=False (Spearman's $r = -0.039$, $p = 0.630$). If the behavior were triggered by an aversion to appearing dishonest, we would expect to observe stronger evidence of subjects considering ENVELOPE_GUESS in the decision to be dishonest in PUBLIC=True. Concerning the effect of preferences for anonymity, we refer to subsection 4.5. We relegate the discussion of the influence of the laboratory and the profits from preceding experiments to Appendix B, since they do not offer any behavioral insights.

4.4. Underclaimers

Contrary to other reporting games which allow for observability (e.g., Abeler et al., 2016; Gneezy et al., 2016), we observe a non-negligible fraction of 7.2% of all participants underclaiming. These are participants who claim less than what they are entitled to and who thus lie to their own disadvantage.

We can only speculate about the reasons for this finding and have not offered any hypotheses earlier, as we did not ex ante expect to find underclaimers. To our knowledge, only Utikal and Fischbacher (2013) address underclaiming. They suggest that some people face such a disproportionally high cost when likely to be perceived as being dishonest that they are led to lie to their disadvantage. Specifically, Utikal and Fischbacher (2013) illustrate this assumption with the results of a die experiment with a group of nuns who appeared to lie to their financial disadvantage. In the context of our experiment, identified underclaimers could consequently be participants with a very high preference for *appearing* honest and a comparably low preference for honesty itself (as underclaiming is, after all, also a form of lying). The hypothetical floor of reports at the honest claim thus would not apply to them. Support for this theory comes from the observation that, while underclaiming occurs in every treatment, underclaimers make up a significantly higher proportion of the participants with ENVELOPE_CONTENT=30 than of those with ENVELOPE_CONTENT=70 (12.3% vs. 1.9%, Fisher's exact test $p < 0.001$). This is in line with our argument for H9 in the hypotheses section 3.3: participants with a strong wish to appear honest choose their final claim with an eye to the maximum possible claim, leading more participants to underclaim with envelope content 30 than with envelope content 70. Based on these findings, however, we would have to conclude that the preference for appearing honest is unrelated to subjects' proneness to feeling shame and to their pro-social preferences, as the correlation of RELATIVE_DEVIATION with both SHAME_SCORE and SVO_ANGLE among underclaimers is very low (Spearman's $r = 0.067$ and $r = 0.013$ respectively, $p > 0.1$ for both).

Concerning these potential explanations, note that 7.2% underclaimers are substantial compared to similar experiments, but with 23 participants nonetheless low in absolute terms. We are thus cautious about the statistical relevance of our findings and explanations. What we can report is that the overall results are not affected by the underclaimers, since the results of our key econometric analysis do not change when we exclude them (see Appendix E for a discussion of observation exclusion rules).

4.5. Preferences for anonymity

In our experiment participants are exogenously assigned to a treatment. Outside of the lab (e.g., when shopping for P2P insurance), people are often able to choose between contracts with anonymity-like characteristics and contracts without. We investigate this choice by asking participants about their preferences after they have received their payoff and have seen the consequences arising from the presence or absence of anonymity. Specifically, we elicit whether or not they would prefer a situation of transparency about participants' reports (including their own) in a reporting situation such as the one they just experienced. We find that their response ANONYMITY_PREFERENCE obtains a marginally significant coefficient value when added to our model as a control variable ($b = 0.41$, $p = 0.076$ for ordinal and $b = 14.01$, $p = 0.018$ for Tobit in Table B.5). Participants who prefer anonymity thus submit higher claims. Figure 7 shows that this observation is independent of whether or not participants actually experienced anonymity or not.

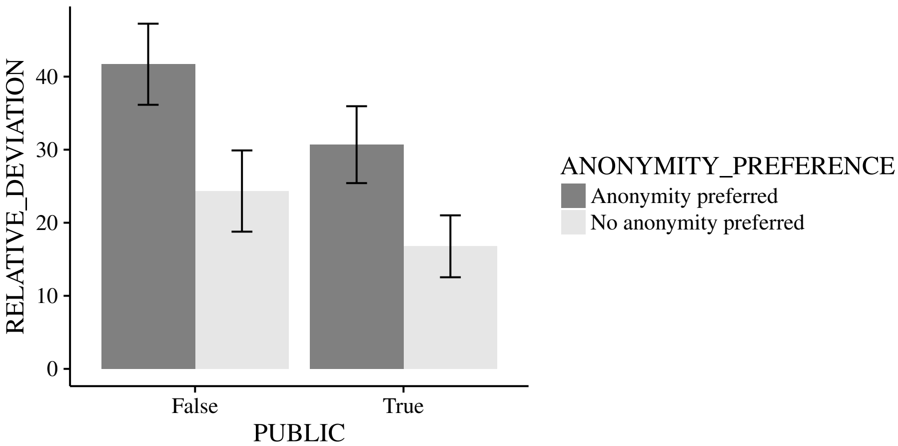

Figure 7: RELATIVE_DEVIATION depending on ANONYMITY_PREFERENCE and experience of PUBLIC. Error bars indicate standard errors.

Unsurprisingly then, as shown in Figure 8, a significant majority of those who report honestly prefer no anonymity (proportions test against 50% with $\chi^2 = 4.93$, $p = 0.026$). Conversely, a majority of the over- and underclaimers prefers anonymity. However, the difference from 50% is only significant for overclaimers, possibly owing to the smaller sample of underclaimers (proportions test against 50%: $\chi^2 = 4.60$, $p = 0.032$ for overclaimers; $\chi^2 = 0$, $p = 1$ for underclaimers). We report this finding as our final result.

Result 7 *Dishonest subjects prefer anonymity while honest subjects prefer transparency.*

This result indicates that schemes incorporating a reduction in experienced anonymity for the reporters should allow for voluntary selection. In this case, more honest reporters will self-select into the scheme, leaving competing schemes to face the downside of adverse selection.

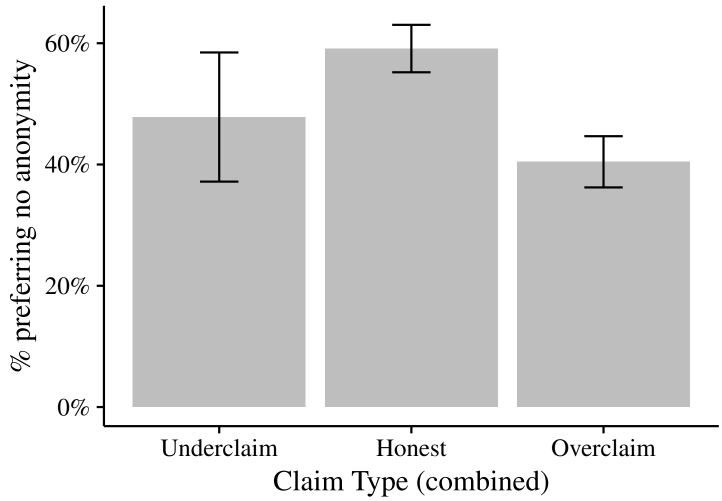

Figure 8: Preference for anonymity by combined claim type (Under-/Overclaim contain both Full UC/OC and Partial UC/OC). Error bars indicate standard errors.

5. Discussion and conclusion

We experimentally confirm the often assumed positive effect on honesty when unverified claims are made from a common pool in which all participants have transparency about each registered claim and about the impact of each claim on the common pool. However, contrary to common reasoning, this effect is exclusively driven by lifting the anonymity of reporters and their claims in our experiment (even when there is no common pool and observers of the claim thus have no direct stake in the reports). We find strong support for a general preference for appearing honest as the main explanation for this effect. Contrary to our initial belief, anticipated shame does not seem to play a role. Any potential positive effects of being concerned about others' payoffs in the externality treatments seem to be offset by participants who fear to lose out when others overclaim, suggesting a positive correlation between fear of exploitation and deception. In total, adding externalities without lifting anonymity of reporters does not seem to significantly affect honesty.

While we believe our study to be highly instructive about the identified effects within its context, we see additional aspects which could inspire future research. First, repeated decision-making and experience could change behavior. For example, Diekmann et al. (2015) and Rauhut (2013) report changes in honesty towards behavior of others in a multi-period die experiment when participants learn about the other participants' (dishonest) behavior.

There are also indications for an erosion of compliance over time, both in public goods games, where explicit minimum contribution rules cannot sustain cooperation over time (Galbiati and Vertova, 2008), and in the honesty literature, where ego-depletion through refraining from lying can also cause more subsequent lying (Gino et al., 2011). Second, some real situations also include the possibility of punishment based solely on perceived honesty and not on actual honesty (e.g., exclusion from a micro-insurance collective). Such a means could very well have an even stronger reinforcing effect than lifting anonymity on its own, as it does in the public goods literature (Fehr and Gächter, 2000). Third, allowing for self-selection into voluntary schemes could lead to different results. While we believe, based on our findings of preferences for anonymity, that the effect is likely to be positive (e.g., more honest participants select into a transparent scheme without anonymity), we would be interested in seeing further investigation into this direction.

Finally, we believe that our findings offer clear advice for practitioners: when trying to increase honesty and compliance in situations affecting common resources, it is more important to focus on putting participants' reports under clear scrutiny by observers than it is to make the impact of reports transparent to the claimant. We expect, e.g., that public disclosure of individuals' taxes paid, or of claims registered for (social) insurance or public subsidies positively affect reporting honesty in the respective situations.

Acknowledgement

We thank James Tremewan and participants of the Austrian Experimental Economics Workshop 2017 for their valuable comments, as well as Nicolas Rüsch for providing the German version of the TOSCA-3 questionnaire. This work received financial support from the University of Graz, which is gratefully acknowledged.

References

Abbink, K., Irlenbusch, B., Renner, E., 2002. An experimental bribery game. Journal of Law, Economics, and Organization 18 (2), 428–454.

Abeler, J., Raymond, C., Nosenzo, D., 2016. Preferences for Truth-Telling. IZA Discussion Paper 10188.

Alm, J., 2012. Measuring, explaining, and controlling tax evasion: Lessons from theory, experiments, and field studies. International Tax and Public Finance 19 (1), 54–77.

Barr, A., Serra, D., 2009. The effects of externalities and framing on bribery in a petty corruption experiment. Experimental Economics 12 (4), 488–503.

Battigalli, P., Charness, G., Dufwenberg, M., 2013. Deception: The role of guilt. Journal of Economic Behavior and Organization 93, 227–232.

Biener, C., Eling, M., Landmann, A., Pradhan, S., 2016. Can Group Incentives Alleviate Moral Hazard? The Role of Pro-Social Preferences. Working Paper, 1–32.

Bø, E. E., Slemrod, J., Thoresen, T. O., 2015. Taxes on the internet: Deterrence effects of public disclosure. American Economic Journal: Economic Policy 7 (1), 36–62.

Bock, O., Nicklisch, A., Baetge, I., 2012. hroot: Hamburg registration and organization online tool. WiSo-HH Working Paper Series 1, 1–14.

Casal, S., Mittone, L., 2016. Social esteem versus social stigma: The role of anonymity in an income reporting game. Journal of Economic Behavior and Organization 124, 55–66.

Conrads, J., Lotz, S., 2015. The effect of communication channels on dishonest behavior. Journal of Behavioral and Experimental Economics 58, 88–93.

Coricelli, G., Rusconi, E., Villeval, M. C., 2014. Tax evasion and emotions: An empirical test of re-integrative shaming theory. Journal of Economic Psychology 40, 49–61.

Crosetto, P., Weisel, O., Winter, F., 2012. A flexible z-Tree implementation of the Social Value Orientation Slider Measure (Murphy et al. 2011). Jena Economic Research Papers 062, 1–8.

Diekmann, A., Przepiorka, W., Rauhut, H., 2015. Lifting the veil of ignorance: An experiment on the contagiousness of norm violations. Rationality and Society 27 (3), 309–333.

Dohmen, T., Falk, A., Huffman, D., Sunde, U., Schupp, J., Wagner, G., 2011. Individual risk attitudes: Measurement, determinants, and behavioral consequences. Journal of the European Economic Association 9 (3), 522–550.

Dufwenberg, M., Dufwenberg, M., 2016. Lies in Disguise A Theoretical Analysis of Cheating. CESifo Working Paper 6208.

Fehr, E., Gächter, S., 2000. Cooperation and Punishment in Public Goods Experiments. The American Economic Review 90 (4), 980–994.

Fischbacher, U., 2007. z-Tree : Zurich toolbox for ready-made economic experiments. Experimental Economics 10, 171–178.

Fischbacher, U., Föllmi-Heusi, F., 2013. Lies in disguise-an experimental study on cheating. Journal of the European Economic Association 11 (3), 525–547.

Fischbacher, U., Gächter, S., 2010. Social Preferences, Beliefs, and the Dynamics of Free Riding in Public Goods Experiments. The American Economic Review 100 (1), 541–556.

Galbiati, R., Vertova, P., 2008. Obligations and cooperative behaviour in public good games. Games and Economic Behavior 64 (1), 146–170.

Gibson, R., Tanner, C., Wagner, A. F., 2013. Preferences for Truthfulness: Heterogeneity among and within Individuals. American Economic Review 103 (1), 532–548.

Gino, F., Schweitzer, M. E., Mead, N. L., Ariely, D., 2011. Unable to resist temptation: How self-control depletion promotes unethical behavior. Organizational Behavior and Human Decision Processes 115 (2), 191–203.

Gneezy, U., 2005. Deception: The Role of Consequences. The American Economic Review 95 (1), 384–394.

Gneezy, U., Kajackaite, A., Sobel, J., 2016. Lying Aversion and the Size of the Lie. Working Paper.

Gneezy, U., Rockenbach, B., Serra-Garcia, M., 2013. Measuring lying aversion. Journal of Economic Behavior and Organization 93, 293–300.

Greenberg, A. E., Smeets, P., Zhurakhovska, L., 2015. Lying, Guilt, and Shame. Working Paper.

Greiner, B., 2015. Subject pool recruitment procedures: Organizing experiments with ORSEE. Journal of the Economic Science Association 1, 114–125.

Grosch, K., Rau, H. A., 2017. Gender differences in honesty: The role of social value orientation. CEGE Discussion Papers 308.

Hao, L., Houser, D., 2017. Perceptions, intentions, and cheating. Journal of Economic Behavior and Organization 133, 52–73.

Hilbig, B. E., Hessler, C. M., 2013. What lies beneath: How the distance between truth and lie drives dishonesty. Journal of Experimental Social Psychology 49 (2), 263–266.

Hurkens, S., Kartik, N., 2009. Would I lie to you? On social preferences and lying aversion. Experimental Economics 12 (2), 180–192.

Irlenbusch, B., Villeval, M. C., 2015. Behavioral ethics: how psychology influenced economics and how economics might inform psychology? Current Opinion in Psychology 6, 87–92.

Jacobsen, C., Fosgaard, T. R., Pascual-Ezama, D., 2017. Why Do We Lie? A Practical Guide To the Dishonesty Literature. Journal of Economic Surveys , forthcom, 1–31.

Joireman, J. A., Lasane, T. P., Bennett, J., Richards, D., Solaimani, S., 2001. Integrating social value orientation and the consideration of future consequences within the extended norm activation model of proenvironmental behaviour. British Journal of Social Psychology 40 (1), 133–155.

Khalmetski, K., Sliwka, D., 2017. Disguising Lies Image Concerns and Partial Lying in Cheating Games. CESIFO Working Paper 6347.

Kleinlercher, D., Stoeckl, T., 2017. On the provision of incentives in finance experiments. Working paper,

under second-round review at Experimental Economics.
Köneke, V., Müller-Peters, H., Fetchenhauer, D., 2015. Versicherungsbetrug verstehen und verhindern. Springer Gabler.
Masclet, D., Noussair, C. N., Tucker, S., Villeval, M.-C., 2003. Monetary and Nonmonetary Punishment in the Voluntary Contributions Mechanism. American Economic Review 93 (1), 366–380.
Mazar, N., Amir, O., Ariely, D., 2008. The Dishonesty of Honest People: A Theory of Self-Concept Maintenance. Journal of Marketing Research 45 (6), 633–644.
Murphy, R. O., Ackermann, K. A., 2013. Social value orientation: Theoretical and measurement issues in the study of social preferences. Personality and Social Psychology Review 18 (1), 1–29.
Murphy, R. O., Ackermann, K. A., Handgraaf, M. J. J., 2011. Measuring Social Value Orientation. Judgment and Decision Making 6 (8), 771–781.
Rasmußen, A., Leopold-Wildburger, U., 2014. Honesty in intra-organizational reporting. Journal of Business Economics 84 (7), 929–958.
Rauhut, H., 2013. Beliefs about lying and spreading of dishonesty: Undetected lies and their constructive and destructive social dynamics in dice experiments. PLoS ONE 8 (11), 1–8.
Rege, M., Telle, K., 2004. The impact of social approval and framing on cooperation in public good situations. Journal of Public Economics 88, 1625–1644.
Rosenbaum, S. M., Billinger, S., Stieglitz, N., 2014. Let's be honest: A review of experimental evidence of honesty and truth-telling. Journal of Economic Psychology 45, 181–196.
Rüsch, N., Corrigan, P. W., Bohus, M., Jacob, G. A., Brueck, R., Lieb, K., 2007. Measuring shame and guilt by self-report questionnaires: A validation study. Psychiatry Research 150 (3), 313–325.
Smithson, M., Verkuilen, J., 2006. A better lemon squeezer? Maximum-likelihood regression with beta-distributed dependent variables. Psychological Methods 11 (1), 54–71.
Steinel, W., De Dreu, C. K. W., 2004. Social motives and strategic misrepresentation in social decision making. Journal of Personality and Social Psychology 86 (3), 419–434.
Tangney, J., Stuewig, J., Mashek, D., 2007. Moral emotions and moral behaviour. Annual Review of Psychology 58, 345–72.
Tangney, J. P., Dearing, R. L., Wagner, P., Gramzow, R., 2000. The test of self-conscious affect-3 (TOSCA-3). Fairfax, VA: George Mason University.
Utikal, V., Fischbacher, U., 2013. Disadvantageous lies in individual decisions. Journal of Economic Behavior and Organization 85, 108–111.
van de Ven, J., Villeval, M. C., 2015. Dishonesty under scrutiny. Journal of the Economic Science Association 1 (1), 86–99.
Van Lange, P. a. M., Bekkers, R., Schuyt, T. N. M., Vugt, M. V., 2007. From Games to Giving: Social Value Orientation Predicts Donations to Noble Causes. Basic and Applied Social Psychology 29 (4), 375–384.
Visher, T., Dohmen, T., Falk, A., Huffman, D., Schupp, J., Sunde, U., Wagner, G. G., 2013. Validating an ultra-short survey measure of patience. Economics Letters 120 (2), 142–145.

Appendix A. Experimental instructions

In this appendix, we provide the translations of the instructions given to the participants. All experiments were run in Austria, therefore the instructions were written in German. The original instructions are available on request.

Appendix A.1. Claim game instructions

Instructions for the *claim game* differ by treatment type. For treatments *Public*, *Ext* and *PublicExt*, participants were made aware that they are in groups of four. Participants in externality treatments were further informed about the specific payoff (i.e., that they would receive an equal share of the remainder in the pool). For participants in the no-anonymity treatments, the last paragraph was added to explain the information other participants receive about their reports. The four texts are displayed in a 2×2 format in table A.4 with comparable paragraphs situated next to each other.

	EXTERNALITY=False	**EXTERNALITY=True**
PUBLIC=False	Every participant in this experiment faces the same task. Each participant is entitled to a total payoff of 1.00 euro. This 1.00 euro stems from two sources. (1) First, there is an envelope on your desk. This envelope contains a cash amount of between 0.00 and 1.00 euros (in 10-cent steps). The money in your envelope belongs to you. The amounts in the envelopes of different participants may differ – not every participant's envelope holds the same amount. Please look into your envelope now. You may pocket the contents of your envelope. You do not have to show these contents to any other participant, nor to the experimenter. During the entire experiment, no other participant will learn the contents of your envelope. (2) In order to obtain the difference towards the total payoff of 1.00 euro you are entitled to, you will enter a request on the next screen. The amount you request may be between 0.00 and 1.00 euros (in 10-cent steps) and will not be verified. At the end of the experiment the experimenter will pay out to you the amount you requested.	You were randomly allocated to a group of four participants for this experiment. The three other participants in your group face the same task as you do. Each participant is entitled to a total payoff of 1.00 euro. This 1.00 euro stems from two sources. (1) First, there is an envelope on your desk. This envelope contains a cash amount of between 0.00 and 1.00 euros (in 10-cent steps). The money in your envelope belongs to you. The amounts in the envelopes of different participants may differ – not every participant's envelope holds the same amount. Please look into your envelope now. You may pocket the contents of your envelope. You do not have to show these contents to any other participant, nor to the experimenter. During the entire experiment, no other participant will learn the contents of your envelope. (2) In order to obtain the difference towards the total payoff of 1.00 euro you are entitled to, you will enter a request on the next screen. The experimenter has prepared an account of 4.00 euros to fulfill the requests of your group. The amount you request may be between 0.00 and 1.00 euros (in 10-cent steps) and will not be checked. At the end of the experiment the experimenter will pay out to you the amount you requested. Money left over in the account after the requests of you and the other three group members have been satisfied will be divided into equal shares for you and the other three group members and will also be paid out.
PUBLIC=True	You were randomly allocated to a group of four participants for this experiment. The three other participants in your group face the same task as you do. Each participant is entitled to a total payoff of 1.00 euro. This 1.00 euro stems from two sources. (1) First, there is an envelope on your desk. This envelope contains a cash amount of between 0.00 and 1.00 euros (in 10-cent steps). The money in your envelope belongs to you. The amounts in the envelopes of different participants may differ – not every participant's envelope holds the same amount. Please look into your envelope now. You may pocket the contents of your envelope. You do not have to show these contents to any other participant, nor to the experimenter. During the entire experiment, no other participant will learn the contents of your envelope. (2) In order to obtain the difference towards the total payoff of 1.00 euro you are entitled to, you will enter a request on the next screen. The amount you request may be between 0.00 and 1.00 euros (in 10-cent steps) and will not be checked. At the end of the experiment the experimenter will pay out to you the amount you requested. After you have entered your request, this request, together with your photo, will be shown to the other three group members. You will also see the photos of the other group members and the request each of them made. This step does not affect the payment of your request.	You were randomly allocated to a group of four participants for this experiment. The three other participants in your group face the same task as you do. Each participant is entitled to a total payoff of 1.00 euro. This 1.00 euro stems from two sources. (1) First, there is an envelope on your desk. This envelope contains a cash amount of between 0.00 and 1.00 euros (in 10-cent steps). The money in your envelope belongs to you. The amounts in the envelopes of different participants may differ – not every participant's envelope holds the same amount. Please look into your envelope now. You may pocket the contents of your envelope. You do not have to show these contents to any other participant, nor to the experimenter. During the entire experiment, no other participant will learn the contents of your envelope. (2) In order to obtain the difference towards the total payoff of 1.00 euro you are entitled to, you will enter a request on the next screen. The experimenter has prepared an account of 4.00 euros to fulfill the requests of your group. The amount you request may be between 0.00 and 1.00 euros (in 10-cent steps) and will not be checked. At the end of the experiment the experimenter will pay out to you the amount you requested. Money left over in the account after the requests of you and the other three group members have been satisfied will be divided into equal shares for you and the other three group members and will also be paid out. After you have entered your request, this request, together with your photo, will be shown to the other three group members. You will also see the photos of the other group members and the request each of them made. This step does not affect the payment of your request.

Table A.4: Instructions of all treatments comprised of PUBLIC and/or EXTERNALITY.

Appendix A.2. SVO slider measure instructions

We decided to write our own instruction to the SVO slider measure, given that there was no suitable German introduction available. Nonetheless, they are very similar to the original instructions in Murphy et al. (2011). We made sure (and advised participants about the fact) that the other participant with which they would share money and from whom they would receive money would be two different people and would not have been in their group in the claim game. The translation of our instructions is:

"Every participant in this experiment faces the same task. You will decide how to allocate certain amounts of euro-cents between yourself and another person. We will refer to this other person simply as "the other" forthwith. The other is a person who you do not know and you will both remain anonymous to each other. All your decisions are entirely confidential, just like the decisions of the other. For each of the following questions, please enter the allocation of money you like best. Your decisions will generate money for you and for the other. One of your distribution decisions will be randomly selected to be paid out. You will be matched with another participant who receives the amount you chose to give to the 'other' in the selected question. You will receive the amount that you allocated to yourself in the selected question.

You will also be the 'other' for another participant and therefore receive money from the decision of this other person. Your 'other' and the participant for whom you are the 'other' are different persons. If you were assigned to a group in previous parts of the final experiment, neither participant was in the same group with you.

There are no right or wrong answers in this task, this is only about your personal preferences. When you have made your decision, mark the corresponding position and click OK. You are only able to mark one position per question. Your decisions influence both the amount of money you receive and the amount of money the other person receives."

Appendix B. Comparison of regression models including controls

In addition to our key independent variables from the econometric model in section 4.2, we collect other data from the questionnaire which may influence honesty in our experiment.

First, we ran the experiment in two different labs, and all session were run as addenda to other experiments. We therefore control for differences in the laboratories using a dummy variable GRAZ (coded as 0 for Innsbruck and 1 for Graz). We also include a variable PREVIOUS_PROFIT which contains the profits obtained in the experiments preceding the claim game, in euros. Second, we collect a range of demographic information, including gender, which may play a part in determining honesty behavior (Abeler et al., 2016). We control for SOCIAL_STATUS_PARENTS (participants self-assess the perceived social status of their parents on a scale from 0=low to 10=high), FEMALE (coded as 0 for male and 1 for female) and AGE (in years). Third, we ask for a self-assessment of RISK_AVERSION following Dohmen et al. (2011), and of PATIENCE following Visher et al. (2013) (both coded from 0=low to 10=high). Fourth, we include questions about participants' beliefs. In ENVELOPE_GUESS we asked participants – after they had entered their claim but before they had seen the payoff – what amount they believed the envelopes in the claim game contained on average. Answers could range from 0 to 100 cents. In ANONYMITY_PREFERENCE we ask participants – after they had claimed and seen their payoff – if, in the situation just experienced, they would prefer anonymity about reporters and reports or not (coded as 0 for no anonymity preferred and 1 for anonymity preferred).

Four coefficients are at least marginally significant. ENVELOPE_GUESS and ANONYMITY_PREFERENCE were already discussed in section 4. Concerning the differences in lab and previous profit, we find a significant influence of these variables. Nonetheless, they do not change our main results substantially. While we cannot know for certain why these factors turn out to be significant, we have strong hypotheses leading us to suspect that the reasons lie outside our treatment manipulations. First, the subjects in Innsbruck have far more experience in participating in experiments than those in Graz (while we cannot assemble the data specifically for our experimental sessions, the average experience, as of April 19, 2017, of subject pool members in Innsbruck is 6.08 sessions vs. 0.99 sessions in Graz). Research by Kleinlercher and Stoeckl (2017) shows that subjects' motivations for participating in experiments are quite diverse when inexperienced (i.e., the first time they participate in an experiment), with only 30.6% naming money as their main motive. This proportion increases to more than 61% for subjects with at least 5 sessions of experience, suggesting that more experienced subjects may be more willing to maximize payoffs than to refrain from doing so in order to remain honest. Second, having greater experimental experience, subjects in Innsbruck are also more likely than subjects in Graz to already have participated in an honesty experiment. Here, Fischbacher and Föllmi-Heusi (2013) show that repeated participation in their die task increases reports yielding higher payoffs. Third, while PREVIOUS_PROFIT is significant, it would be cause for worry only if the coefficient was negative (e.g., a lower payoff in the previous experiment leading to higher dishonesty). This could signify that participants want to make good on unfavorable previous outcomes. As this is not the case, we rather believe in the following channel: participants eager to make money in general (as displayed in the preceding experiment) are also more likely to be dishonest in the claim game in order to maximize payoffs. This indicates consistent behavior instead of being cause for concern.

To check the stability of our results, we also apply alternative regression models. Besides ordinal regression for studying the levels of CLAIM_TYPE, and Tobit regression for studying RELATIVE_DEVIATION, we conduct an ordinary least squares regression to study RELATIVE_DEVIATION. After transforming RELATIVE_DEVIATION into the (0,1)-interval,[9] we also conduct a beta regression, once with a constant precision parameter and once with a precision parameter depending on our key variables. This allows us to fit more complex relationships than just linear ones.

The results for our main regressors are highly consistent across models in terms of direction and significance. While we get some additional significant controls in individual models (like PATIENCE or AGE in the beta regression), we do not consider them to be of great relevance or reliability.

[9]As proposed by Smithson and Verkuilen (2006).

	Ordinal	Tobit	OLS	Transformed Beta/const	Transformed Beta/Var
INTERCEPT		117.06^{***}	97.04^{***}	2.67^{***}	2.61^{***}
		(33.21)	(26.93)	(0.70)	(0.66)
PUBLIC	-0.94^{***}	-21.70^{***}	-15.34^{**}	-0.52^{***}	-0.43^{***}
	(0.32)	(8.31)	(6.67)	(0.17)	(0.17)
EXTERNALITY	-0.36	-6.29	-4.26	-0.07	-0.23
	(0.33)	(8.31)	(6.64)	(0.17)	(0.19)
PUBLIC x EXTERNALITY	0.79^{*}	16.55	12.36	0.30	0.40
	(0.46)	(11.72)	(9.44)	(0.25)	(0.25)
SHAME_SCORE	-0.04^{**}	-1.01^{**}	-0.84^{**}	-0.02^{**}	-0.02^{**}
	(0.02)	(0.44)	(0.35)	(0.01)	(0.01)
GUILT_SCORE	0.01	0.26	0.27	0.01	0.01
	(0.02)	(0.60)	(0.48)	(0.01)	(0.01)
SVO_ANGLE	-0.04^{***}	-0.99^{***}	-0.77^{***}	-0.02^{***}	-0.01^{***}
	(0.01)	(0.23)	(0.18)	(0.00)	(0.00)
ENVELOPE_CONTENT	1.05^{***}	21.98^{***}	19.77^{***}	0.39^{***}	0.43^{***}
	(0.26)	(6.68)	(5.39)	(0.14)	(0.13)
GRAZ	-0.97^{***}	-24.76^{***}	-19.56^{***}	-0.57^{***}	-0.56^{***}
	(0.25)	(6.32)	(5.11)	(0.13)	(0.13)
PREVIOUS_PROFIT	0.04^{***}	0.88^{**}	0.69^{**}	0.02^{**}	0.02^{***}
	(0.01)	(0.39)	(0.31)	(0.01)	(0.01)
RISK_AVERSION	0.02	1.42	0.91	0.04	0.02
	(0.05)	(1.32)	(1.06)	(0.03)	(0.03)
PATIENCE	-0.08^{*}	-1.55	-1.05	-0.04^{*}	-0.04^{**}
	(0.04)	(1.05)	(0.85)	(0.02)	(0.02)
SOCIAL_STATUS_PARENTS	0.02	0.31	0.28	-0.02	-0.02
	(0.07)	(1.91)	(1.53)	(0.04)	(0.04)
FEMALE	-0.26	-7.86	-5.05	-0.15	-0.14
	(0.25)	(6.44)	(5.20)	(0.14)	(0.13)
AGE	-0.03	-0.83	-0.84	-0.02	-0.02^{**}
	(0.03)	(0.65)	(0.53)	(0.01)	(0.01)
ENVELOPE_GUESS	-0.02^{**}	-0.47^{*}	-0.44^{**}	-0.01	-0.01
	(0.01)	(0.26)	(0.21)	(0.01)	(0.01)
ANONYMITY_PREFERENCE	0.41^{*}	14.01^{**}	10.88^{**}	0.31^{**}	0.24^{**}
	(0.23)	(5.93)	(4.81)	(0.13)	(0.11)
Precision: INTERCEPT				0.83^{***}	0.13
				(0.07)	(0.58)
Precision: PUBLIC					0.66^{***}
					(0.20)
Precision: EXTERNALITY					-0.49^{**}
					(0.20)
Precision: SHAME_SCORE					0.03^{***}
					(0.01)
Precision: GUILT_SCORE					-0.02
					(0.01)
Precision: SVO_ANGLE					0.03^{***}
					(0.01)
Precision: ENVELOPE_CONTENT					0.01
					(0.14)
Precision: PUBLIC x EXTERNALITY					0.14
					(0.28)
AIC	712.82	2808.19			
BIC	788.06	2875.90			
Log Likelihood	-336.41	-1386.09		164.90	189.87
Deviance	672.82	390.20			
Num. obs.	318		318	318	318
Total		318			
Left-censored		3			
Uncensored		247			
Right-censored		68			
Wald Test		107.52			
R^2			0.27		
Adj. R^2			0.23		
RMSE			41.03		
Pseudo R^2				0.26	0.24

$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

Table B.5: Comparison of different regression models: Ordinal regression studying CLAIM_TYPE as defined in section 4; Tobit regression studying RELATIVE_DEVIATION, left censored at -100 and right censored at 100; ordinary least squares regression studying RELATIVE_DEVIATION; beta regressions studying RELATIVE_DEVIATION transformed by $((\text{RELATIVE\_DEVIATION} + 100)/200 * (N-1) + 0.5)/N$, as proposed by Smithson and Verkuilen (2006).

Appendix C. Comparison of DEVIATION and RELATIVE_DEVIATION

To make sure that RELATIVE_DEVIATION can be pooled across ENVELOPE_CONTENT, we compare distributions of the measure across this dimension. Figure C.9 displays the split and combined histograms of both DEVIATION and RELATIVE_DEVIATION, as well as fitted density functions. While it is not surprising that the distributions for DEVIATION differ significantly between envelope contents (Kolmogorov-Smirnov test $D = 0.366$, $p < 0.001$), we find no substantial difference when using RELATIVE_DEVIATION (Kolmogorov-Smirnov test $D = 0.129$, $p = 0.140$).

Table C.6 compares Tobit regressions using DEVIATION and RELATIVE_DEVIATION as the dependent variables. It does not document any substantial differences when interpreting the results, which supports our use of RELATIVE_DEVIATION as a suitable measure.

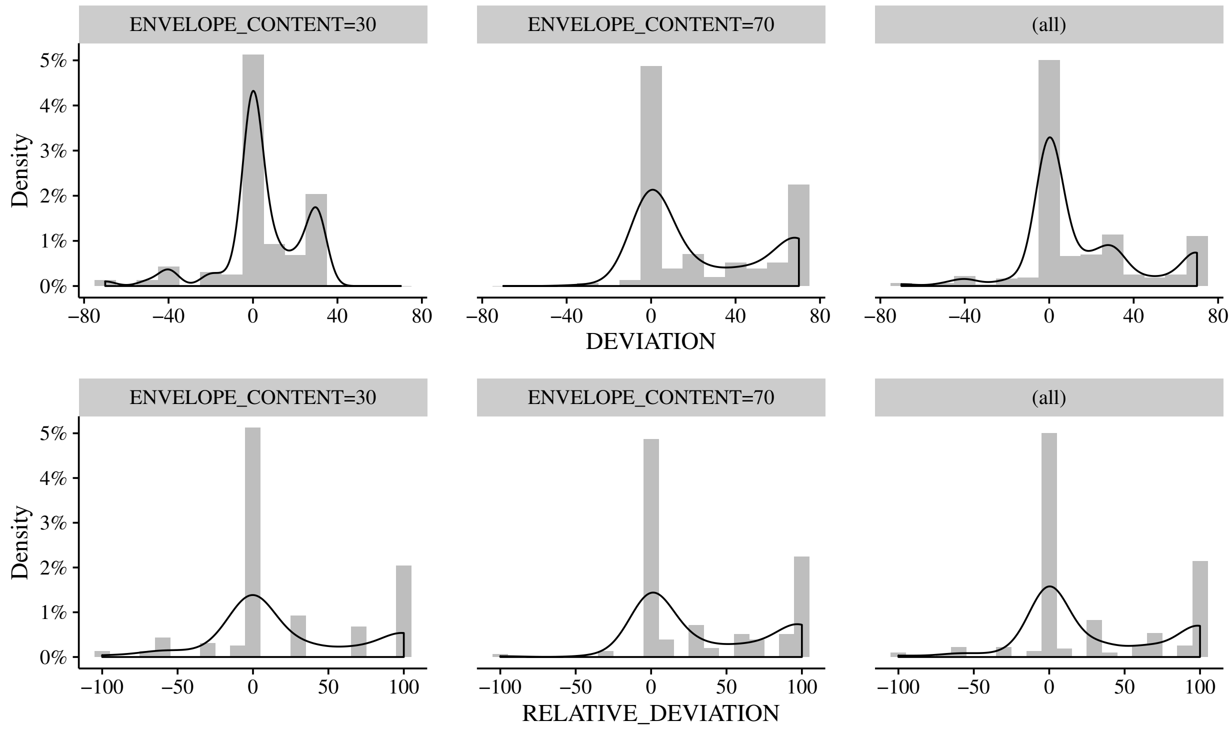

Figure C.9: Histograms and fitted non-parametric densities of DEVIATION and RELATIVE_DEVIATION, by ENVELOPE_CONTENT and combined (all).

	DEVIATION	RELATIVE_DEVIATION
INTERCEPT	43.04^{***}	117.06^{***}
	(16.42)	(33.21)
PUBLIC	-8.66^{**}	-21.70^{***}
	(4.08)	(8.31)
EXTERNALITY	-4.20	-6.29
	(4.06)	(8.31)
PUBLIC x EXTERNALITY	8.60	16.55
	(5.76)	(11.72)
SHAME_SCORE	-0.64^{***}	-1.01^{**}
	(0.22)	(0.44)
GUILT_SCORE	0.26	0.26
	(0.30)	(0.60)
SVO_ANGLE	-0.37^{***}	-0.99^{***}
	(0.11)	(0.23)
ENVELOPE_CONTENT	26.92^{***}	21.98^{***}
	(3.29)	(6.68)
GRAZ	-11.52^{***}	-24.76^{***}
	(3.11)	(6.32)
PREVIOUS_PROFIT	0.42^{**}	0.88^{**}
	(0.19)	(0.39)
RISK_AVERSION	0.53	1.42
	(0.65)	(1.32)
PATIENCE	-0.64	-1.55
	(0.52)	(1.05)
SOCIAL_STATUS_PARENTS	0.28	0.31
	(0.94)	(1.91)
FEMALE	-4.59	-7.86
	(3.16)	(6.44)
AGE	-0.42	-0.83
	(0.32)	(0.65)
ENVELOPE_GUESS	-0.22^{*}	-0.47^{*}
	(0.13)	(0.26)
ANONYMITY_PREFERENCE	7.30^{**}	14.01^{**}
	(2.92)	(5.93)
AIC	2724.34	2808.19
BIC	2792.06	2875.90
Log Likelihood	-1344.17	-1386.09
Deviance	368.31	390.20
Total	318	318
Left-censored	2	3
Uncensored	281	247
Right-censored	35	68
Wald Test	151.89	107.52

$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

Table C.6: Comparison of Tobit regressions of DEVIATION (left/right censored at -70/70) and of RELATIVE_DEVIATION (left/right censored at -100/100) including controls.

Appendix D. TOSCA-3 and SVO slider results

Figure D.10 reports box plots of the most relevant scales we elicited directly (SHAME_SCORE, GUILT_SCORE, SVO_ANGLE) per treatment. The figure documents that they are properly spread across the full scale. We also find no differences between treatments (ANOVA, $p \gg 0.1$ for each scale).

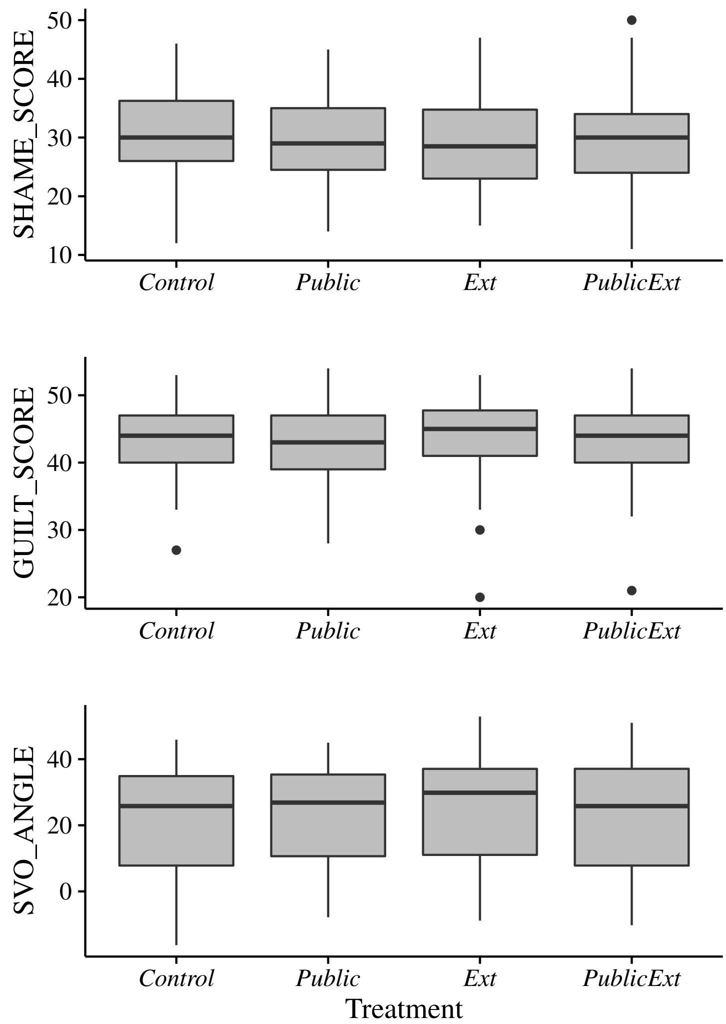

Figure D.10: Box plots of TOSCA-3 and SVO results by treatment.

Appendix E. Participant exclusion rules

We exclude participants from our dataset who clearly indicated (in the questionnaire free text field at the end of the experiment) or were observed making an error (e.g., not opening the envelope at all). This led to the exclusion of 10 participants (or 3% of total) and the dataset henceforth referred to as "Used". In this section, we compare results of "Used" to the full dataset "Full" (no exclusions), as well as to a very selective dataset "Selective". For "Selective" we start from "Used" and in addition exclude underclaimers (as underclaims were unexpected and could have been caused by erroneous reading of the instructions) as well as participants in three individual sessions from Graz which had a different preceding experiment than the others. This leads to the exclusion of a total of 66 participants or 20.1% of the total. As can be seen from the interaction plots in Figure E.11, overall claims increase, in "Selective", mainly due to the exclusion of the underclaimers. However, the regression results do not change substantially for our key variables of interest. This gives us additional confidence that our data selection "Used" does not create spurious causality findings.

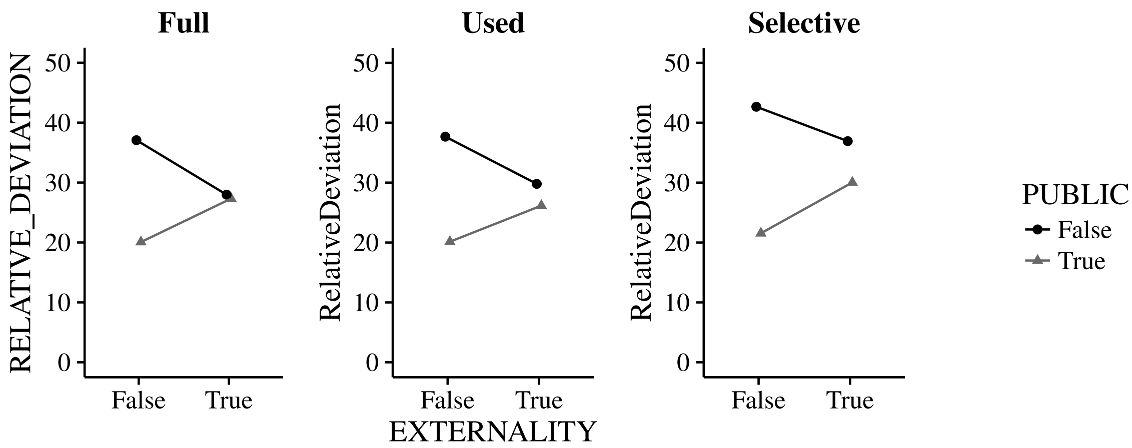

Figure E.11: Interaction plots by data selection.

	Full	Used	Selective
INTERCEPT	110.95***	117.06***	120.38***
	(33.36)	(33.21)	(35.56)
PUBLIC	−21.00**	−21.70***	−28.47***
	(8.27)	(8.31)	(8.25)
EXTERNALITY	−8.17	−6.29	−4.76
	(8.26)	(8.31)	(8.14)
PUBLIC x EXTERNALITY	19.05	16.55	19.75*
	(11.70)	(11.72)	(11.84)
SHAME_SCORE	−0.98**	−1.01**	−0.92**
	(0.44)	(0.44)	(0.43)
GUILT_SCORE	0.26	0.26	0.26
	(0.60)	(0.60)	(0.60)
SVO_ANGLE	−0.98***	−0.99***	−1.11***
	(0.23)	(0.23)	(0.22)
ENVELOPE_CONTENT	21.51***	21.98***	12.35*
	(6.62)	(6.68)	(6.87)
GRAZ	−26.70***	−24.76***	−26.48***
	(6.32)	(6.32)	(6.21)
PREVIOUS_PROFIT	0.82**	0.88**	−0.10
	(0.39)	(0.39)	(0.72)
RISK_AVERSION	1.33	1.42	2.65*
	(1.33)	(1.32)	(1.38)
PATIENCE	−1.29	−1.55	−1.50
	(1.05)	(1.05)	(1.04)
SOCIAL_STATUS_PARENTS	−0.03	0.31	−1.10
	(1.88)	(1.91)	(1.87)
FEMALE	−7.61	−7.86	−8.03
	(6.46)	(6.44)	(6.44)
AGE	−0.65	−0.83	0.01
	(0.65)	(0.65)	(0.71)
ENVELOPE_GUESS	−0.37	−0.47*	−0.38
	(0.26)	(0.26)	(0.27)
ANONYMITY_PREFERENCE	12.00**	14.01**	8.04
	(5.90)	(5.93)	(6.11)
AIC	2904.32	2808.19	2246.68
BIC	2972.60	2875.90	2310.91
Log Likelihood	-1434.16	-1386.09	-1105.34
Deviance	403.14	390.20	321.53
Total	328	318	262
Left-censored	3	3	0
Uncensored	255	247	200
Right-censored	70	68	62
Wald Test	104.70	107.52	101.88

$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

Table E.7: Tobit regressions studying RELATIVE_DEVIATION by dataset, left censored at -100 and right censored at 100.