CESifo WORKING PAPERS

10897
2024

January 2024

The Truth-Telling of Truth-Seekers: Evidence from Online Experiments with Scientists

Moritz A. Drupp, Menusch Khadjavi, Rudi Voss

CESifo

Impressum:

CESifo Working Papers
ISSN 2364-1428 (electronic version)
Publisher and distributor: Munich Society for the Promotion of Economic Research - CESifo GmbH
The international platform of Ludwigs-Maximilians University's Center for Economic Studies and the ifo Institute
Poschingerstr. 5, 81679 Munich, Germany
Telephone +49 (0)89 2180-2740, Telefax +49 (0)89 2180-17845, email office@cesifo.de
Editor: Clemens Fuest
https://www.cesifo.org/en/wp
An electronic version of the paper may be downloaded
· from the SSRN website: www.SSRN.com
· from the RePEc website: www.RePEc.org
· from the CESifo website: https://www.cesifo.org/en/wp

CESifo *Working Paper No. 10897*

The Truth-Telling of Truth-Seekers: Evidence from Online Experiments with Scientists

Abstract

Academic honesty is crucial for scientific advancement, yet replication crises and misconduct scandals are omnipresent. We provide evidence on scientists' truth-telling from two incentivized coin-tossing experiments with more than 1,300 scientists. Experiment I, with predominantly European and North-American scientists, shows that fewer scientists over-report winning tosses when their professional identity is salient. The global Experiment II yields heterogeneous effects. We replicate Experiment I's effect for North-American scientists, but find the opposite for Southern European and East-Asian scientists. Over-reporting correlates with publication metrics and country-level measures of academic and field-experimental dishonesty, suggesting that country-level honesty norms also guide truth-telling by scientists.

JEL-Codes: C930, D820, K420, J450.

Keywords: truth-telling, lying, identity, science, cross-country, experiment.

Moritz A. Drupp
University of Hamburg / Germany
Moritz.Drupp@uni-hamburg.de

Menusch Khadjavi
Free University Amsterdam / The Netherlands
m.khadjavipour@vu.nl

Rudi Voss
University of Kiel / Germany
voss@economics.uni-kiel.de

This version: January 13, 2024
We are grateful to Alain Cohn, John A. List, Michel Maréchal, Sarah Necker, Martin Quaas, Marie-Catherine Riekhof and Jörn Schmidt as well as audiences at Amsterdam, Bonn, Halle, Kassel, Maastricht, Marburg, NYU, UCSB, UCSD and Vienna for helpful discussions. We thank the participating scientists, the administrative office of the science organization for providing e-mail address data, Jörn Schmidt for handling correspondences, Marie-Catherine Riekhof for helping with the study design, Olaf Bock and his team at the Experimental Lab in Hamburg for support in administering the experiments, as well as Pia Förster, Clara Paczkowski and Lasse Renz-Kiefel for research assistance. At the time of conducting the experiments, no IRB was available at the respective institutions. The Research Laboratory of the Faculty of Business, Economics and Social Sciences at the University of Hamburg, Germany, confirmed that our experiment is in accordance with its ethical code of conduct. This work was supported by the German Ministry of Education and Research [grant 01UT1410] and by the Cluster of Excellence 80, funded by the German Research Foundation (DFG).

1. Introduction

Whether and to what degree scientists behave ethically sound and tell the truth is of fundamental importance for the development of science, for public trust in science and, as such, for the future of mankind. Marshall (2000: 1162) called this “a Million-Dollar Question”, but this number is likely a gross underestimate. This is particularly true for times which call, on the one hand, for more “evidence-based policy-making” and have been otherwise guided by a tendency to blur distinctions between objective knowledge and so-called “alternative facts”, “fake news” and ‘post-truths”.

The Merriam-Webster Dictionary (2017) defines science as “knowledge or a system of knowledge covering general truths or the operation of general laws especially as obtained and tested through scientific method”. The quest for ensuring integrity in research conduct is probably as old as science itself, yet the reputation of *truthful* science has in particular suffered in recent times from prominent instances of scientific misconduct.[1] A famous, now retracted, article by Wakefield et al. (1998) suggested that vaccinating children against measles, mumps and rubella increases their risk of autism. Poland and Jacobson (2011) describe the public reaction of anti-vaccination campaigns to the now disproved article. In the time following the publication of Wakefield et al. (1998), there was a record of hundreds of cases of measles outbreaks, causing a number of children to die (Poland and Jacobson, 2011), providing some indication of the tremendous social costs of scientific misconduct. The long-term costs of this anti-vaccination case of dishonest science became apparent during the global COVID-19 pandemic.

Beyond the prominent cases of scientific misconduct mentioned above, survey evidence suggests that a considerable number of scientists engage in a broader set of questionable research practices (see, for example, John et al., 2012; List et al., 2001; Martinson et al., 2005; Necker, 2014).[2] A meta-study by Fanelli (2009) summarizes findings from 21 individual studies and shows that around two percent of scientists admit to having

---

[1] These include i.a. cases such as of the cloning expert Hwang Woo-suk, the evolutionary biologist Marc Hauser and social psychologist Diederik Stapel. Articles by Sang-Hun (2009), Wade (2010) and Bhattacharjee (2013) provide more information on the respective misconduct. More recently, research on dishonesty itself has become subject to investigations on data fabrication (O’Grady 2021; The New Yorker 2023).

[2] Besides anonymous survey-based approaches, there are a number of other recent examples testing research integrity and the robustness of scientific research: For example, Camerer et al. (2016) run replications of 18 experimental economic studies and find that about two-thirds of all findings can be replicated; Brodeur et al. (2016) provide evidence that the reporting of empirical findings tends to be biased towards specifications that favor rejecting the null hypothesis, while Brodeur et al. (2020) examine substantial variation in p-hacking and publication bias across causal inference methods. To improve research practices, Simmons et al. (2011) recently proposed rules of sound scientific conduct in order to decrease so-called experimenter degrees of freedom. Blanco-Perez and Brodeur (2020) find that one such approach—editorial statements encouraging null results—increases the proportion of reported null results findings by double digit percentage points.

committed serious forms of scientific misconduct at least once, such as fabricating, falsifying or modifying data or results. It further supports findings of a study by Martinson et al. (2005) showing that as many as one-third of scientists admit to have engaged in questionable research practices, such as "using another's ideas without obtaining permission or giving due credit", "failing to present data that contradict one's own previous research", or "inappropriately assigning authorship credit". This literature suggests that the search for general truths is not always conducted in a truthful manner. Yet, this evidence only relies on anonymous survey responses, subject to the challenge that there is no individual (monetary) incentive to participate and to report truthfully.

Our study provides experimental economic evidence on incentivized truth-telling of more than 1,300 scientists by means of two online (field) experiments. We thus provide evidence that can be viewed as complementary to above mentioned survey approaches. Specifically, our aim is to investigate whether the professional identity as a scientist motivates and fosters truthful behavior and to study how truth-telling varies across scientific contexts—across disciplines, countries as well as academic status, as proxied by number of publications and citations. We build in particular on two studies by Cohn et al. (2014, 2015), who provide experimental evidence that bankers and prisoners behave less honestly when their respective professional identity is made salient as compared to a (private) control identity (cf. Villeval, 2014). Subsequent research by Rahwan et al. (2019) yields qualitatively similar but insignificant effects of priming bankers on dishonest reporting and suggests that effects may vary across countries, among other things due to different banking cultures. They further find that priming non-bankers to think about their professional context does not have a significant effect on dishonest reporting. In contrast, we hypothesize that the norms and behavioral patterns associated with working as a scientist may imply the opposite, i.e. greater truth-telling in the professional context. After all, science "consists in the search for truth" (Popper, 1996).

To this end, we employ a simple coin-tossing task in which scientists are asked to toss a fair coin four times and report back their number of tail tosses (Abeler et al., 2014). For each reported tail toss they receive five Euros. While individual (dis)honesty is not detectable, we can estimate the deviation of reported tosses against the expected truthful distribution. Studying truth-telling in this manner has become a major research focus in economics.[3] Furthermore, a substantial number of studies show that such a task carries

[3] For instance, see Abeler et al. (2014, 2019), Cappelen et al. (2013), Cohn et al. (2014, 2015), Fischbacher and Föllmi-Heusi (2013), Gächter and Schulz (2016), Gibson et al. (2013), Gneezy (2005), Gneezy et al.

external validity as it correlates with truth-telling behavior beyond the simple experimental task (e.g., Cohn et al., 2015; Cohn and Maréchal, 2018; Dai et al., 2018; Drupp et al., 2019; Gächter and Schulz, 2016; Potters and Stoop, 2016).

To study whether professional identity of scientists induces more honesty, we draw on the identity priming literature that was developed in social psychology and is now an active research field within economics (see Cohn and Maréchal (2016) for a recent review).[4] The idea is that individuals have multiple identities that are guided by different norms and behavioral patterns (Akerlof and Kranton, 2000). Individuals experience disutility if they deviate from norms prescribed by their respective salient identity.

Our experimental design consists of two treatments. The professional identity treatment aims at making a participant's professional identity as a scientist salient, while the private identity (control) treatment aims at making the private identity salient. To prime participants, we use nine simple questions that are designed to capture common features of a professional or private context, unrelated to truth-telling and as similar as possible across the two treatments. For example, participants in the professional identity treatment were asked "Where did you last go to for a conference/workshop?" and "What activity in your work do you enjoy the most?", while participants in the private identity (control) treatment were asked "Where did you last go on holiday?" and "What activity in your leisure time do you enjoy the most?". In the context of our study, the priming intervention aims to reveal the behavioral difference between a participant's private and professional identity and thus be indicative of the norms and behavioral patterns associated with the scientist identity of the participants in terms of truth-telling and honesty.

We collected data in two online experiments with scientists. For Experiment I we were able to get access to the mailing list of an international scientific organization concerned with marine science to invite scientists. We conducted this experiment in summer 2016. Experiment II was a follow-up motivated by calls to test the replicability and generalizability of empirical research (e.g., Maniadis et al. 2017). For this, we pre-registered our main hypotheses[5] based on the evidence from Experiment I and expanded

(2013, 2018), Houser et al. (2016), Mazar et al. (2008), Pasqual-Ezama et al. (2015), Potters and Stoop (2016), Rosenbaum et al. (2014).

[4] As our study concerns the technique of priming and focuses on truth-telling behavior, it is worthwhile to note that there are doubts about the robustness of results obtained in the priming literature in social psychology and suspicions that questionable research practices have been employed. As a response to this critique, Daniel Kahneman called for systematic replication efforts in this field (Young, 2012). Not specifically scrutinizing priming studies, Camerer et al. (2016) and Open Science Collaboration (2015) have recently demonstrated that such large-scale replication attempts are feasible and fruitful.

[5] Available on the OSF platform under https://osf.io/vkf62.

the scope in terms of the number of observations, the global distribution of participants and the diversity in terms of academic disciplines. Experiment II ran in spring 2019.

In Experiment I, including 437 responses to our coin-tossing task from predominantly European and North American marine scientists, we find that fewer scientists over-report winning coin tosses in the professional identity treatment compared to the private identity treatment, suggesting that the identity as a scientist may indeed entail stronger honesty norms that induce more truth-telling. We further relate over-reporting in the experiment with publication metrics of individual respondents and find that lower reporting of winning tail tosses is associated with higher measures of scientific output or scientific success (proxied by having published in *Science, Nature* or *PNAS*).

The data from Experiment II, including 864 observations, fails to replicate our main result from Experiment I on aggregate and reveals substantial heterogeneity in the treatment effects between world regions and disciplines. While we cannot reject the null hypothesis of honest reporting for scientists from Northern and Eastern Europe, which makes treatment effects for these world regions infeasible, we replicate our initial hypothesis, supported in Experiment I, for North American scientists. Scientists from other world regions significantly over-report winning tail tosses compared to the expected truthful frequency and treatment effects even point into the opposite direction for Southern European and Eastern Asian scientists. While honesty norms associated with scientific identity thus can principally increase truth-telling, the prevalent norm needs to be truthful behavior in the first place. As a plausibility check for external validity, we further correlate our tail toss reporting measure—on the country level—with two other country-level measures of dishonesty: first, the lost wallet measure of Cohn et al. (2019) and, second, the scientific misconduct measure of Ataie-Ashtiani (2018). Both dishonesty measures correlate with ours in an expected manner: Higher tail toss reporting is associated with fewer returned wallets in Cohn et al. (2019) and with a higher scientific misconduct ranking in Ataie-Ashtiani (2018), suggesting that cross-cultural differences in truth-telling also carry over to reporting practices of scientists.

Our findings underscore the importance of variations in honesty norms and the context-dependence of professional cultures across geographies and suggest caution in generalizing effects from European and North American samples (cf., Rahwan et al. 2019). In conclusion, relying on honesty norms for science across the globe appears ineffective and it is crucial to establish rigorous measures for preventing scientific misbehavior to ensure that science is not derailed from its path to generate truths.

2. Experiment I

In sub-section 2.1 we describe the experimental design, procedures and our identity-economic hypotheses. We report results from Experiment I in sub-section 2.2.

2.1 Experimental Design

To study the truth-telling of scientists, we conducted an online (field) experiment with members of an international scientific organization that was established more than 100 years ago.[6] The administrative office of the organization provided an e-mail list of its 1,930 members. In the summer of 2016 we contacted all members by e-mail and invited them to participate in a short online study that consisted of ten pages and took about 15 minutes to complete. We told them that they could earn €25 on average (equivalent to $27 at the time of the experiment) for participating, with the exact individual earnings depending on chance and their choices. We ensured that their individual responses are kept confidential and informed the participants about the confidentiality.

Upon clicking the link to the online study in the invitation e-mail, participants were assigned to one of two treatments by the online platform: either the professional identity treatment (abbreviated *Professional* or *PROF*) or the private identity (control) treatment (*Private* or *PRIV*). A preamble page provided further details on the experiment and the mode of payment (Amazon vouchers). The study then began with simple descriptive questions on age, gender and nationality. This was followed by our manipulation that consisted of nine questions either relating to their professional identity (*Professional* treatment) or relating to their private identity (*Private* treatment). The purpose of these questions was to make the participants' professional identity as scientists, and associated norms, more salient in *Professional* as compared to *Private*.

The behavioral intervention of identity priming builds on a by now established strain of the experimental economics literature.[7] The basic idea—based on Akerlof and Kranton (2000)—is that people have multiple identities that are guided by different norms

[6] The members are predominantly natural scientists, with a focus on the marine environment. We do not report the name of the scientific organization in our paper to increase the anonymity of our respondents. Upon request we are open to provide more information for academic transparency, of course.

[7] Cohn and Maréchal (2016) provide a review of identity priming in economics and discuss how this builds on a previous substantial literature in social psychology. The first economic experiments on identity priming were Chen and Li (2009) as well as Benjamin et al. (2010). There are two general approaches to studying how behavioral measures differ across identities: (1) artificially inducing certain identities or (2) studying the effect of identity priming in natural populations, such as bankers (Cohn et al., 2014), criminals, (Cohn et al., 2015), or scientists, as in our study.

and behavioral patterns. Individuals experience disutility if they deviate from norms prescribed by their respective salient identity. This depends on the relative weight of that identity. The technique of identity priming aims at making a given identity, such as the professional identity of being a scientist, temporarily more salient (see, e.g., Benjamin et al., 2010, 2016; Cohn and Maréchal, 2016, Cohn et al., 2014, 2015, 2018).

Our study design closely builds on the approach of Cohn et al. (2014, 2018). The priming intervention should reveal the behavioral difference between a participant's private and professional identity. Thus, the intervention should be indicative of the norms and behavior associated with the scientific identity as compared to the private identity of the participants in terms of truth-telling and honesty. In an effort to reduce potential confounding due to priming effects that are unrelated to their private or professional identity, we designed the questions to capture salient features of their professional work or private life identity, yet to be as similar as possible in terms of their content and context. For example, participants in the professional treatment were asked "Where did you last go to for a conference/workshop?", while participants in the private control treatment were asked "Where did you last go on holiday?" (see Table 1 for a list of all priming questions posed and Appendix A for screenshots from the online survey). These priming questions were the only difference between the two treatment conditions.[8]

Table 1: Identity priming questions

***Professional* identity treatment**	***Private* identity treatment**
Who is your current employer?	What is your current city of residence?
How many years have you worked for this institution?	How many years have you lived in your current accommodation?
Do you have a tenured position?	Are you married?
How large is your direct working team (yourself included)?	How large is your direct family (yourself included)?
Where did you last go to for a conference/workshop?	Where did you last go on holiday?
In which year did you start your PhD?	In which year did you kiss the first boy/girl?
At what time do you usually arrive at the office?	At what time do you usually arrive at home?
What activity in your work do you enjoy the most?	What activity in your leisure time do you enjoy the most?
How satisfied are you with your work in general?	How satisfied are you with your life in general?

[8] The only other difference was that on the preamble page we stated that the study was on either on "Work [Life] satisfaction, including individual attitudes and behavior" in *Professional* [*Private*].

This identity manipulation was followed by three experimental tasks. First, participants were asked to complete a risk preference elicitation task based on Binswanger (1981) and Eckel and Grossman (2002), the results of which we analyze in a companion paper (Drupp et al., 2020). The risk task was followed by the truth-telling task based on Abeler et al. (2014) that is the main focus of this paper. We present this task in more detail below. Finally, we posed a hypothetical social time preference task. The three tasks were always presented in this order and it was not possible to switch back once a participant had proceeded to the next page. The lottery outcome of the risk task was only revealed at the end of the experiment and thus could not have affected coin toss reporting.

Following the experimental tasks, participants were also asked to complete a short follow-up survey that included a word-completion task designed to provide an implicit measure of how well the identity priming manipulation had worked (cf. Cohn et al., 2014). Participants were presented with eight-word fragments and they were asked to fill in the gaps with letters to form existing words. The idea is that when the professional identity is salient other words come to the participants' mind as compared to when the private identity is salient. For example, they were shown the word fragment "j o u r_ _ _", which they could complete with the word "journal" that scientists would frequently encounter in their professional lives, or the word "journey", which might be more salient to those in the *Private* treatment.[9] We classified all completed words and either assigned the number 1 to words related to the professional work identity or number 0 to words classified as related to a private life. Words that could not be classified as relating to either context or words without actual meaning were coded as missing.[10] Together with the payoff from the risk elicitation task, ranging from 2 to €16, and a €5 compensation for completing the short follow-up survey, each participant could earn up to €41.[11] The payoff from the risk task was revealed after participants had completed the follow-up survey. Finally, we offered the possibility to donate fractions of the earnings to the charity "Doctors Without Borders".[12]

For studying the truth-telling of scientists, we adapt the 4-coin-tossing task of Abeler et al. (2014) for our online field experiment. Participants were asked to use any coin

---

[9] The first two of the eight-word fragments ("_ a l k" and "_ o o k") had no unambiguous professional science interpretation. These two were meant as an easy start for participants and served, following Cohn et al. (2014, 2017), the purpose of disguising the purpose of the task. The other word fragments were: " _ i s _", "_ _ s s i o n", "c o _", "_ _ o c k" as well as "_ _ p e r".

[10] When in doubt about a word's meaning we relied on the Merriam-Webster dictionary.

[11] The design thus aimed at paying out all participants. Overall, we spent 3,389 Euros on participant remuneration and donated 6,199 Euros to "Doctors Without Borders" on our participants' behalf.

[12] This donation option was not pre-announced and it thus could not have influenced coin toss reporting as it was not possible to move back within the study.

that has the usual "tails" and "heads" format (see Appendix A for a screenshot of the task). The participants' task was then to toss this coin exactly 4 times and report their tail toss result by clicking on the relevant button in a table.[13] For each instance they reported that the winning toss "tails" laid on top, they received €5. An important feature of this task is that lying can be detected only on aggregate when examining the distribution of decisions, but not on the individual level. Thus, depending on chance and honesty, each participant received between 0 and €20 for this task. Similar experiments using coin tosses or dice rolling have been conducted to answer a whole range of related research question. Abeler et al. (2019) provide a meta-study on truth-telling behavior summarizing results based on 72 individual studies. Several key insights emerge from this burgeoning literature: (i) Participants only over-report on average a quarter of the possible maximum pay-off and thus exhibit substantial lying costs; (ii) Participants' reporting behavior is not significantly influenced by stake sizes; (iii) female participants over-report somewhat less compared to males; (iv) students over-report more than other participants. Testing different models that can be used to explain reporting behavior, Abeler et al. (2019) find that models that combine a preference for being honest, i.e. that entail a utility cost for deviating from the truthful response, and preference for being seen as honest, i.e. that entail individual reputation concerns, perform best in explaining experimental data.[14]

As our main contribution is not a focus on modeling lying costs but more directly on the effect of making the professional scientific identity more salient vis-a-vis the private identity, we follow Benjamin et al. (2010) and Cohn et al. (2015) in relying on a simple behavioral choice model that features the salience of distinct identities. The model of reporting behavior considers an overall lying aversion due to deviating from the truthful response that may differ between the two identities, which may be guided by different norms and behavioral patterns.[15]

In absence of a possibility to detect individual lying, an individual i faces a trade-off between monetary incentives and (moral) costs of lying. While the individual derives utility only from her payoff proportional to the reported number of coin tosses r_i, she also

[13] As we could not ensure the availability of coins to toss remotely, we offered the option to proceed without reporting one of the five tail toss possibilities in case they could not organize a coin to toss. They were told that they would not receive a payoff for this task in this case. No participant clicked this option.

[14] Another recent study by Gneezy et al. (2018) investigates how lying costs depend on the size of the lie in various dimensions using both unobservable as well as observable lying tasks. Besides intrinsic lying costs considered in our model, they find that reputational concerns can drive honest reporting.

[15] Besides the application of identity-priming model to truth-telling behavior of criminals by Cohn et al. (2015), this model has been employed for explaining effects of religious identity on a suite of economic preferences (Benjamin et al., 2016) and on risk preferences (Cohn et al., 2017; Drupp et al., 2020).

suffers disutility from reporting a number that deviates from the true number of tail tosses, r_{it}. The individual payoff-maximizing choice is given by r_{ip}. Aggregating over all n individuals of a population yields the mean tail toss reporting $\bar{R} = \frac{1}{n}\sum_{i=1}^{n} r_i$, which can be disaggregated for different groups within a population. For instance, we denote the mean tail toss reporting in the *Professional* identity treatment as $\bar{R}^{PROF}$.[16]

Furthermore, let $\hat{R}^{PROF}$ ($\hat{R}^{PRIV}$) denote the expected reporting behavior implied by prevailing norms in the professional environment (private identity context). In the context of our study, these norms imply certain lying costs, with $\hat{R} = \frac{\lambda}{2}\ (r_i - r_{it})$, where λ is a parameter determining the degree of overall lying aversion.[17] As the degree of lying aversion may depend on expected behavior and prevailing norms in different contexts, it may in particular differ across the private and the professional identity conditions, i.e. $\lambda^{PROF} \neq \lambda^{PRIV}$ and thus $\hat{R}^{PROF} \neq \hat{R}^{PRIV}$. Furthermore, let s denote the strength of the identification with the professional environment. Let $w_i(s) \in [0,1]$ denote how much weight the individual puts on complying with expectations in the professional environment, which depends on the strength of identifying with the respective environment, with $\frac{\partial w_i}{\partial s} \geq 0$. In this set-up, the individual chooses her reporting r_i to maximize utility

$$\max_{r_i}\ U_i(r_i) \ = \ -\tfrac{1}{2}\left(1 - w_i(s)\right)\left(r_i - \ \hat{R}^{PRIV}\right)^2 - \tfrac{1}{2} w_i(s)\left(r_i - \hat{R}^{PROF}\right)^2. \qquad (1)$$

The optimal tail toss reporting r_i^* is a weighted average of the "expected" reportings under both identities,

$$r_i^* = \left(1 - w_i(s)\right)\hat{R}^{PRIV} + w_i(s)\hat{R}^{PROF} \ . \qquad (2)$$

In terms of the model, our priming experiment aims at varying the salience of the professional or the private identity and thus the strength s of identifying with the professional identity. Priming participants with the professional identity (the *Professional* treatment) should increase s, while priming the private identity (the *Private* treatment) should decrease s. Participants should therefore (weakly) experience an increase in the weight put on one identity or the other when completing our experimental task. As such,

[16] While the model considers continuous reporting, our subsequent experiment is based on a setting where possible reporting levels are discrete, with $r_i, r_{it} \in \{0{,}4\}$. Furthermore, the mean truthful response is given by $R_t = \frac{1}{n}\sum_{i=1}^{n} r_{it} = 2$, and the payoff-maximizing choice is given by $R_p = \frac{1}{n}\sum_{i=1}^{n} r_{ip} = 4$.

[17] While conceptually both over- and under-reporting weigh equally strongly, empirical evidence suggests that under-reporting is negligible. For instance, Gneezy et al. (2018) find that only one out of 602 participants under-reports to his or her disadvantage.

the treatment effect should reveal the marginal behavioral impact of the primed identity and its associated norms relative to the other treatment,

$$\frac{\partial r_i^*}{\partial s} = \frac{\partial w_i}{\partial s}(\hat{R}^{PROF} - \hat{R}^{PRIV}). \tag{3}$$

Based on previous findings in the experimental literature (Abeler et al., 2019), we expect heterogeneity regarding individual truth-telling r_i^* in our sample of scientists. Translating the average standardized estimate of the meta-study of Abeler et al. (2019) into our context predicts an average tail toss report $\bar{R}$ of 2.44. We formulate:

Hypothesis 1: *Average over-reporting is in-between the truthful and the payoff maximizing choice.*

While previous research has shown that professional identity is associated with higher over-reporting of winning coin tosses (i.e. lower truth-telling) for bankers and criminals (Cohn et al., 2014, 2015), we hypothesize that the norms and behavioral patterns associated with working as a scientist implies greater truth-telling. After all, science is a system of knowledge covering general truths (Popper, 1996). We therefore assume greater lying costs in the professional science context, $\lambda^{PROF} > \lambda^{PRIV}$, and accordingly norms associated with lower expected mean tail toss reporting, $\hat{R}^{PROF} < \hat{R}^{PRIV}$. Our model thus predicts that $\frac{\partial r_i^*}{\partial s} < 0$, summarized as

Hypothesis 2: *Average over-reporting of scientists is lower in the professional identity treatment.*

Even though we expect that stronger honesty norms are present in the professional scientific as compared to the average private context, the accumulating evidence on the use of questionable research practices among scientists suggests that we should not expect truthful reporting on average even in the professional identity treatment. For example, if one-third of scientists would lie partially by over-reporting one tail-step, as the anonymous survey evidence cited above might suggest, we would expect an average tail toss reporting of 2.31 tails, leading to

Hypothesis 3: *Even in the professional identity treatment, average reporting behavior differs from the truthful distribution.*

As part of a comprehensive analysis of truth-telling behavior of scientists in the next section, we will confront these hypotheses with our experimental data.

2.2 *Results*

We have received 599 responses to the survey, amounting to a response rate of more than 30%.[18] 437 responses contain a coin toss report. Participants come predominantly from Europe and North America. There are 58% male participants in our sample. The mean age of our participants is 43 years and 52% of our participants have a tenured position.

Before we turn to scrutinizing the decisions in the coin-tossing task, we test whether our implicit measure of identity priming using the word completion task indicates that priming has been successful. For each participant, we aggregate over the given numbers assigned to completed words for the six potential word checks (1 for words associated with professional life, 0 for words associated with private life) and compare the mean value of these aggregate numbers for the two treatments. We find that the mean number of "professional" words, such as "journal", "paper" or "session", is with 2.89 higher in *Professional* as compared to the 2.66 "professional' words in *Private* (t-test: p=0.053).[19] We therefore find that our *Professional* treatment was able to make the professional scientific identity of our participants more salient compared to the *Private* treatment.

We now examine the coin toss reporting behavior of scientists. Figure 1 shows the theoretical binomial distribution for four tosses of a fair coin (blue dots connected by the dashed line), which is the distribution that we would expect if all participants report the outcome of their four coin tosses truthfully. The probability that four times tossing a coin results in r_{it} = 0 or 4 (1 or 3) [2] times tails is 6.25% (25%) [37.5%]. We refer to this distribution as the "truthful distribution", with a mean truthful response of $\bar{R}_t = 2$ tail tosses. The mean payoff-maximizing choice would be the reporting of $\bar{R}_p = 4$ tail tosses. The colored bars in Figure 1 show actual reporting behavior of the participating scientists across the two treatments: *Private* and *Professional*.

First, we analyze overall coin toss reporting by aggregating results from both treatments. Overall reporting by scientists differs strongly from payoff-maximization, with 2.32 tail tosses on average, indicating substantial lying costs. Yet, we also find that scientist over-report tail tosses to their advantage: A Kolmogorov–Smirnov test for comparing overall reporting behavior against the binomial distribution confirms that scientists over-

[18] Overall, 946 individuals clicked on the link to our study. We dropped 10 observations because they responded more than once and one observation because we could identify her as still being a master student. 162 participants completed some parts of the initial demographic questions, priming questions or the risk task, but did not complete the coin-tossing task.

[19] All p-values reported in this paper are based on two-sided tests.

report tail tosses (p<0.001). We therefore cannot reject Hypothesis 1.

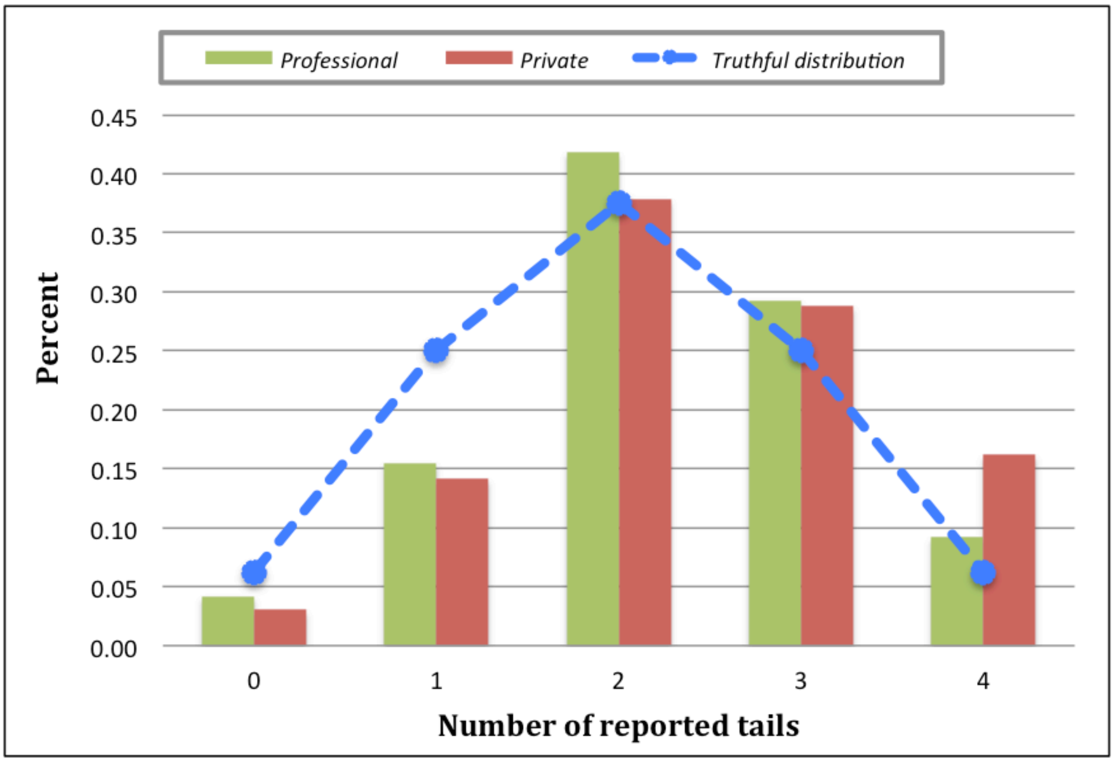


Figure 1: Tail toss-reporting of scientists in the *Private* identity (red bars) and the *Professional* identity treatment (green bars) in Experiment I. The blue, dashed line with dots corresponds to the expected distribution if every scientist reported the true outcomes of their coin tosses. The payoff-maximizing reporting was four times tails.

We now analyze truth-telling in our two treatments. Figure 1 shows reporting behavior of scientists in the private compared to the professional identity treatment. Participants in *Private* report 2.41 tail tosses on average, which is higher than the average report in *Professional* of 2.24 tail tosses (t-test: p=0.073). In particular, we find that scientists in *Professional* report fewer four times tails as compared to those in *Private* (9.21% vs. 16.16%; chi-squared test: p=0.028). This confirms our central Hypothesis 2 and establishes

Result 1: Reporting behavior under professional identity priming

Scientists in the professional identity treatment report, on average, lower tail tosses compared to those in the private identity treatment.

Even though there is fewer over-reporting of higher tail tosses among scientists in *Professional* compared to the *Private* control treatment, we still find that there is over-reporting of tail tosses among those primed with their professional identity: A Kolmogorov–Smirnov test for comparing overall reporting behavior in *Professional* against the expected truthful binomial distribution rejects the null hypothesis at $p<0.01$. That is, the coin-toss reporting in *Professional* still deviates from the truthful distribution, thus confirming Hypothesis 3. Summarizing this finding yields

Result 2: Reporting behavior in the *Professional* identity treatment compared to the truthful distribution

Scientists in the professional identity treatment over-report tail tosses compared to the expected truthful distribution.

As the marginal behavioral impact of increasing the salience of the professional or private identity will depend on the individual baseline salience level (cf. Benjamin et al., 2010), we make use of having inquired about the participant's location when completing the survey to explore differences in reporting behavior across locational contexts.[20] We compare responses of participants who responded from their usual workplace "at work" (n=252) with those being "not_at_work", composed of "at home" as well as "home office" (n=139). We find that the identity priming treatment effect is particularly strong for those scientists responding while not being at their workplace. While the mean number of "professional" words in *Private* is with 2.65 virtually the same as the 2.66 for the whole sample, we find that the mean number of "professional" words in *Professional* for those not at work is 3.11 and thus considerably higher than in *Private* (t-test: p=0.044). While there is no tail toss reporting difference across treatments for scientists responding from their workplace (t-test: p=0.821), the priming intervention had a particularly strong effect on tail toss reporting for those not at their usual workplace: Average tail tosses reported are 2.53 in *Private* and 2.10 in *Professional* (t-test: p=0.008). For four times tails reporting, we find relative frequencies of 18.18% in *Private* and 4.11% in *Professional* (t-test: p=0.007).

Result 3: Identity priming and coin toss reporting effects across locational contexts

The professional identity priming and treatment effect on lower over-reporting is particularly pronounced when participants respond from locations other than their usual workplace.[21]

[20] Pre-offered options were "at work", "at home" and "home office", and a residual "other" option.

[21] Note that as the variables "at_work" and "treatment" are not significantly correlated (t-test: $p > 0.55$); this locational effect does not drive our main treatment effect.

We further relate tail toss reporting to the two other behavioral measures that we collected: risk preferences and donations.[22] First, we elicited risk preferences using the so-called Eckel-Grossman task (Binswanger, 1981; Eckel and Grossman, 2002). Unlike previous studies that examined the relationship between risk-taking and truth-telling,[23] we find that higher tails reporting is associated with higher risk-taking ($p<0.007$).[24] We explore the effects of professional identity priming on risk-taking behavior of scientists in more detail in a companion paper (Drupp et al., 2020). As Drupp et al. (2020) find no significant difference in the overall identity priming treatment effect on risk-taking, we have no indication that the negative correlation between risk-taking and truth-telling is not driving our truth-telling results in this study.

Second, we allowed participants to donate fractions (in 10% steps) of their earnings at the end of the experiment to the NGO "Doctors Without Borders", providing us with an eleven-point step measure of the payoff-fraction donated. This option was not announced earlier, so their donation decision could not have impacted tail toss reporting, but their coin toss reporting and resulting pay-off level might have impacted subsequent donations. We find that participants reporting higher tail tosses are associated with lower step-level donations (correlation-coefficient: -0.17; t-test: $p=0.001$). Indeed, the donation fraction decreases monotonically with reported tail tosses (from 94% for 0 tail tosses to 52% for 4 tail tosses). Yet, we find that the absolute donation amount increases monotonically with reported tail tosses (from €11 for 0 tail tosses to €17 for 4 tail tosses), resulting from higher pay-offs for people with higher reported tail tosses (t-test: $p=0.004$).[25] Furthermore, we find that those who do not donate at all report on average 2.50 tail tosses as compared to only 2.17 tail tosses for those who donate all of their pay-off (t-test: $p=0.009$). Overall, this suggests some consistency of pro-social behavior as revealed by both truth-telling and donation levels and yields

Result 4: Relationship between reporting behavior and donations

Lower over-reporting of tail tosses is, on average, associated with a higher share of subsequent donations.

---

22 Tail toss reporting is not associated with participants' elicited degree of social time preference (t-test: $p>0.70$). The same holds for the year of birth ($p>0.70$), gender ($p>0.90$), being married ($p>0.15$) and having tenure ($p>0.35$), as revealed by two-sided t-tests.

23 For example, Abeler et al. (2014), who rely on a stated preference measure for the German population, or Drupp et al. (2020), who use the same Eckel-Grossman risk-elicitation task.

24 Zimerman et al. (2014) examine the relationship between a stated-preference measure of risk-taking specifically in the domain of ethical risks and find that the stated measure of risk-taking in ethical context is positively correlated with dishonest behavior as elicited using a coin tossing task.

25 We find no difference in fractions donated across *Private* and *Professional* ($p > 0.60$). Also, for the absolute donation amount we find no differences across treatments ($p > 0.35$).

Finally, we relate tail toss reporting to observable measures of scientific output or success: number of citations, h-index and publications as listed in SCOPUS.[26] We find that the number of publications, the h-index and the number of citations are associated with lower over-reporting of winning tail tosses (linear regressions, controlling for age: p=0.079, p=0.084 and p=0.096 respectively).[27] Furthermore, we have gathered data on whether our participants have published in the general science journals *Science, Nature* and *PNAS*. The 36 participants who have published in these journals only report 2.06 tails on average and thus tend to report fewer winning tail tosses (t-test, controlling for age: p=0.106).[28] Figure 2 depicts these findings using linear fit. We therefore conclude that honest reporting tends to be related to having a higher scientific output or success. We summarize

Result 5: Relationship between reporting behavior and scientific output

Lower over-reporting of tail tosses is associated with higher measures of scientific output or success.

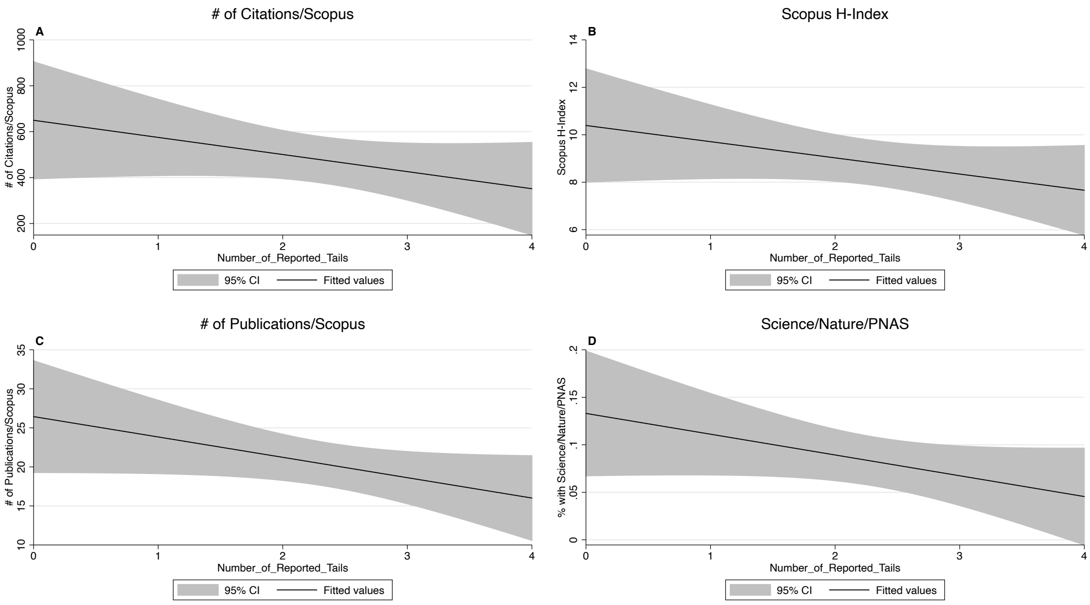


Figure 2: Tail toss-reporting and measures of scientific output gathered in the search engine SCOPUS: The number of citations and publications, the h-index as well as having published in the journals *Science*, *Nature* and *PNAS*.

[26] We find no difference in observable measures of scientific output across *Private* and *Professional* for both number of publications (t-test: p>0.65) and citations (t-test: p>0.85).

[27] Without conditioning on age, the p-values for the number of publications, h-index and the number of citations are p=0.069, p=0.161 and p=0.153, respectively.

[28] Without conditioning on age, the p-value is p=0.100. We have made the conservative assumption that for those for whom we could not obtain information on whether they have published in either of these three journals have not published there. If we disregard those for whom we could not obtain this information, the p-value (when controlling for age) would be p=0.076 (p=0.079).

3. Experiment II

Experiment II broadens the scope of our analysis to world regions beyond North America and Europe and to academic disciplines beyond marine sciences. The purpose of Experiment II is thus to check whether Experiment I's treatment effect replicates and how much heterogeneity we observe across world regions and disciplines. We pre-registered our main hypotheses based on the evidence from Experiment I and expanded the scope in terms of the number of observations, the global distribution of participants and the diversity in terms of academic disciplines. Experiment II ran in March and April 2019.

3.1 Experimental Design

We employed the same between-subjects design as in Experiment I.[29] The procedure of inviting scientists from diverse academic disciplines was operated via the established online search platform SCOPUS that provides corresponding authors' e-mail addresses of publications in peer-reviewed and indexed journals in all scientific disciplines. The platform allowed us to sort the scientists' publications by academic disciplines and to balance the number of observations by discipline and treatment cell. Specifically, we sent out invitation e-mails for participation in our study to a random sample of corresponding authors from eight different scientific subjects, with two subjects from each of the four major science categories life sciences, social sciences, health science and physical sciences, as categorized by SCOPUS. These eight specific scientific subjects are Biochemistry, Genetics and Molecular Biology; Economics, Econometrics and Finance; Environmental Sciences; Medicine; Nursing; Pharmacology, Toxicology and Pharmaceutics; Physics and Astronomy; and Psychology. All corresponding authors have (co-)authored publications included in SCOPUS and were published in 2017. In addition, we invited a random sample of corresponding authors of publications in *Science*, *Nature* and *PNAS* from the year 2017.

For the number of observations, we were limited by our budget of about 30,000 EUR for Experiment II. Given the expected payout of around \$25 per participant and total expenditure of around \$27 per participant, due to an additional donation task, we aimed at collecting a total of 1,080 observations: 432 observations in *Private*, 432

[29] We made two marginal adjustments to the design: First, we added a separate gender treatment to examine effects on risk-taking reported in Drupp et al. (2020) that we do not include here. Second and relatedly, we adjusted the priming questions slightly to separate the *Private* identity treatment from the additional gender treatment (see Table A.1 in Appendix A for details).

observations in *Professional*, plus 216 observations in the gender treatment studied in Drupp et al. (2020).[30] Hence, our complete dataset for Experiment II in this paper includes approx. 48 observations per cell (i.e. per scientific discipline and treatment combination), i.e. 864 observations for our *Professional* and *Private* treatments. In Experiment II's survey question part we asked participants to inform us about the country where they work, so that we could examine geographical variation in the treatment effect.

3.2 Results

Our analysis explores to what extent the professional-identity treatment effect we detected in Experiment I replicates and whether discipline-specific and geographic factors play additional roles. Before we discuss our treatment effects across world regions and disciplines, it is important to examine whether the word completion task yields a similar indication of successful priming by our identity priming questions as in Experiment I.[31] In Experiment II, we only qualitatively find a similar direction of the priming but cannot detect successful priming for the mean number of professional words (t-test: 2.605 vs. 2.502, p=0.232), similar to what Rahwan et al. (2019) find for Middle Eastern Bankers. Exploratory analysis reveals that we only find a successful priming indication for "j o u r n _ _", for which more scientists answered "journal" in *Professional* compared to *Private* (65.9% vs. 50.5%; Chi-squared test: p<0.000). Given the much weaker priming success in Experiment II, any effects we may detect might thus be conservative estimates.

We first provide a general picture of the data by examining the aggregated, average tail-toss reporting. First, scientists over-report tail tosses on average (t-test: p<0.000). Second, average tail-toss reports do not differ noticeably between the *Private* (2.35) and *Professional* (2.38) treatments. Figure 2 depicts the distribution of reporting across treatments. Testing for the treatment effect at this aggregated level for all disciplines and world regions together, a two-sided t-test cannot reject the null hypothesis of equal tail-toss reporting at p=0.733.[32]

---

[30] We sent out invitations at different times of the day, so that different time zones for scientists around the world should not influence participation. Given that one of our word completion tasks contains "Sunday" and "Monday" as solutions, we only sent out invitations on Tuesdays, Wednesdays and Thursdays.

[31] The word completion task in Experiment II included the seven words "_ a l k", "_ _ d a y", "j o u r n _ _", "g r _ _ t", "_ _ s s i o n" and "_ _ p e r" and "_ o o k". Just as in Experiment I, the two words "_ a l k" and "_ o o k" had no unambiguous professional science interpretation and, following Cohn et al. (2014, 2017), were meant to disguise the purpose of the task. For the other five words, we pre-determined word completions that fit either the professional or the private environment of scientists. The responses were coded accordingly and observations with nonsensical and missing completions were dropped for the analysis.

[32] With respect to some of the pre-registered hypotheses and unlike Experiment I, average reported tail tosses in Experiment II do not show any indication for significant interaction effects between the

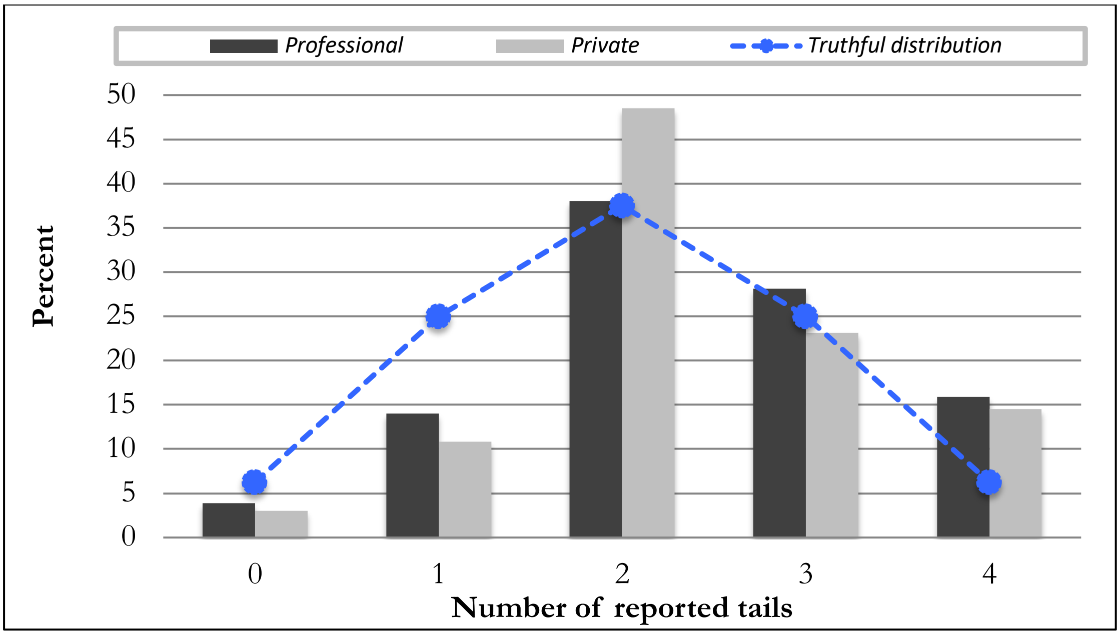


Figure 3: Tail toss-reporting of scientists in the *Private* identity (grey bars) and the *Professional* identity treatment (black bars) in Experiment II. The blue, dashed line with dots corresponds to the expected distribution if every scientist reported the true outcomes of their coin tosses. The payoff-maximizing reporting was four times tails.

Similar to Experiment I, we made an effort to collect measures of scientific output for the scientists in Experiment II from the platform SCOPUS. We were able to identify 498 scientists on SCOPUS and collect the number of citations and number of publications at the moment we ran Experiment II. As pre-registered and similar to Experiment I, we expected negative correlations between reported tail tosses and these output measures (while controlling for age). Figure 4 depicts the two univariate correlations, which are qualitatively consistent with our results from Experiment I. While the regression coefficients are negative and the correlation between citations and tail tosses is significant in a univariate OLS regression (p=0.044), the analysis does not confirm a significant relation once we control for age (p=0.926 for publications, p=0.197 for citations).

Professional treatment and the place from where individuals participated in the experiment. We also find no significant evidence for the pre-registered hypotheses that reported tail tosses positively correlate with the proportion of participating scientists' proportion of studies using data and with the characteristic of scientists who conducted commissioned research for corporations. As pre-registered, we do find that reported tail tosses negatively correlate with scientists' view of scientists as honest and unbiased seekers of the truth.

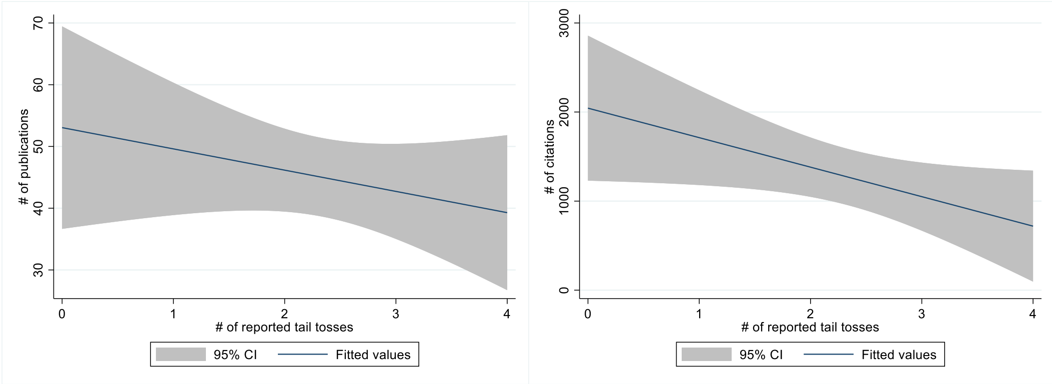


Figure 4: Tail toss-reporting and measures of scientific output gathered in the search engine SCOPUS: the number of publications and citations.

As a key aim of Experiment II is to examine potential heterogeneity of effects across scientific disciplines and world regions, we next analyze the data on disaggregated levels. For the eight scientific disciplines and the *Science*, *Nature* and *PNAS* group we run separate (two-sided) t-tests for the expected difference between *Private* and *Professional* and also test each discipline's mean reporting against the expected true mean of 2 tails. The test statistics of the reported tail-tosses by academic disciplines are summarized in Table 2. Against our hypothesis—but unsurprisingly given the very weak priming indication—we find no significant treatment effects when we split our data by discipline (with the exception of a marginal effect for "Physics and Astronomy" going in the opposite direction). The data, however, reveals level-differences in tail reports between disciplines with the lowest level for "Psychology" and the highest for "Pharmacology, Toxicology and Pharmaceutics". All mean reported tail tosses differ clearly from the truthfully expected number of 2 tails at $p<0.01$, while this deviation from truthful reporting is less significant for the Psychology ($p=0.056$) and the *Science*, *Nature* and *PNAS* groups ($p=0.020$). Figure A.4 in Appendix A depicts the histograms for the different academic disciplines.

Table 2: Mean reported tail-tosses, by treatments and academic disciplines.

Scientific discipline	*Private*	*Professional*	t-test: *Priv* vs *Prof*	t-test: all vs '2'
Biochemistry, Genetics and Molecular Biology	2.31	2.29	p = 0.919	p = 0.003
Economics, Econometrics and Finance	2.54	2.42	p = 0.557	p < 0.001
Environmental Sciences	2.29	2.30	p = 0.974	p = 0.003
Medicine	2.33	2.60	p = 0.133	p < 0.001
Nursing	2.48	2.46	p = 0.919	p < 0.001
Pharmacology, Toxicology and Pharmaceutics	2.49	2.46	p = 0.885	p < 0.001
Physics and Astronomy	2.19	2.53	p = 0.095	p = 0.001
Psychology	2.18	2.23	p = 0.832	p = 0.056
Science, *Nature* and *PNAS*	2.38	2.10	p = 0.182	p = 0.020

Note: 47-49 observations in each cell, 432 total observations per treatment. Two-sided t-tests.

We similarly split our dataset by world regions and test for identity priming treatment effects and differences between the total mean and the expected true mean of 2 tails. As reported in Table 3, the observations per world regions vary between 21 for South Eastern Asia and 161 for Southern Europe.[33] This unequal distribution may not be surprising, given the unequal representation of authors from different world regions in peer-reviewed journals. Most t-tests cannot reject the null hypothesis for treatment differences. There are three exceptions: We find the same treatment effect as in Experiment I for North American scientists (2.35 in *Private* vs. 1.99 in *Professional*, p = 0.031), while we find significant effects in the opposite direction for Southern European scientists (2.16 in *Private* vs. 2.54 in *Professional*, p = 0.0158) and Eastern Asian scientists (2.45 in *Private* vs. 2.94 in *Professional*, p = 0.0811). Testing against the expected truthful mean of 2, we cannot reject the null hypothesis of truthful reporting for Eastern and Northern European scientists (t-tests: p = 0.288 and p = 0.172, respectively). For all other world regions, we detect over-reporting (see Table 3).[34]

[33] Four pre-defined world regions, Caribbean Latin America, Central Asia, Oceania and Central America, are excluded in the analysis as they feature too few observations for meaningful comparisons (2, 1, 11 and 10 observations respectively).
[34] Figure A.5 in Appendix A depicts the histograms for the different world regions.

Result 6: Reporting behavior in world regions compared to the truthful distribution

We detect that scientists, on average, over-report tail tosses, with the notable exceptions of Eastern and Northern European scientists, for whom we cannot reject that they report truthfully.

Table 3: Mean reported tail-tosses, by treatments and world regions.

World Region	***Private***	***Professional***	**t-test: *Priv* vs *Prof***	**t-test: all vs '2'**
Africa (n = 53)	2.57	2.93	p = 0.201	p < 0.001***
Eastern Asia (38)	2.45	2.94	p = 0.081*	p < 0.001***
Eastern Europe (69)	2.12	2.11	p = 0.963	p = 0.288
North. America (145)	2.35	1.99	p = 0.031**	p = 0.061*
North. Europe (50)	2.14	2.17	p = 0.901	p = 0.172
South America (41)	2.12	2.47	p = 0.227	p = 0.086*
South East. Asia (21)	2.66	2.44	p = 0.677	p = 0.036**
Southern Asia (97)	2.58	2.51	p = 0.712	p < 0.001***
South. Europe (161)	2.16	2.54	p = 0.016**	p < 0.001***
Western Asia (33)	2.71	3.06	p = 0.388	p < 0.001***
West. Europe (132)	2.34	2.24	p = 0.517	p < 0.001***

Note: The four pre-defined world regions Caribbean Latin America, Central Asia, Oceania and Central America are excluded in this analysis as they feature too few observations for meaningful comparisons (2, 1, 11 and 10 observations respectively). Two-sided t-tests.

While the means and test statistics in Tables 2 and 3 are aimed at providing a transparent disaggregate picture of our data, they need to be investigated further by means of a regression analysis. We therefore ran ordered logit regressions with discipline- and world region-dummies and additional controls that we collected in a short survey at the end of Experiment II. We report the results of the regressions in Table 4, where we defined North American environmental scientists as the baseline group (motivated by the maritime scientists in Experiment I). The regressions confirm that the *Professional* identity treatment effect of lower tail-toss reporting replicates for this baseline group that is closely related to the sample of Experiment I (p=0.038, p=0.032 and p=0.016 respectively for the three regressions in Table 4). Yet, interaction effects of the *Professional* treatment dummy with other world regions reveal vast heterogeneity of the treatment effect. The regressions confirm that the treatment effect may even affect tail-toss reports in the opposite direction, as indicated by the test statistics in Table 3 for Eastern Asia and Southern Europe.

Table 4: Ordered Logit regression analysis for Experiment II.

	Dependent variable: reported tail tosses		
Independent variables	**(I)**	**(II)**	**(III)**
Professional treatment (dummy)	-0.647** (0.313)	-0.672** (0.313)	-0.749** (0.312)
Age (cont.)	-0.016** (0.007)	-0.016** (0.007)	
Female (dummy)	-0.033 (0.147)	-0.084 (0.151)	
Tenured (dummy)	-0.063 (0.139)	-0.111 (0.142)	
Risk-taking[35] (cont., EG task)	0.081** (0.037)	0.081** (0.037)	
Africa (dummy)	0.404 (0.478)	0.245 (0.485)	0.315 (0.477)
South America (dummy)	-0.602 (0.426)	-0.690 (0.429)	-0.547 (0.419)
Eastern Asia (dummy)	0.108 (0.457)	0.143 (0.464)	0.297 (0.457)
South Eastern Asia (dummy)	0.730 (0.574)	0.573 (0.580)	0.607 (0.583)
Southern Asia (dummy)	0.371 (0.347)	0.281 (0.355)	0.348 (0.349)
Western Asia (dummy)	0.618 (0.511)	0.534 (0.514)	0.734 (0.514)
Eastern Europe (dummy)	-0.565 (0.392)	-0.647 (0.399)	-0.552 (0.394)
Northern Europe (dummy)	-0.376 (0.451)	-0.494 (0.456)	-0.438 (0.454)
Western Europe (dummy)	-0.131 (0.306)	-0.203 (0.309)	-0.122 (0.307)
Southern Europe (dummy)	-0.360 (0.307)	-0.431 (0.312)	-0.412 (0.309)
Africa X *Prof*	1.311** (0.614)	1.235** (0.617)	1.361** (0.614)
South America X *Prof*	1.405** (0.670)	1.399** (0.674)	1.411** (0.668)
Eastern Asia X *Prof*	1.551** (0.652)	1.521** (0.652)	1.530** (0.652)
Western Asia X *Prof*	1.627** (0.745)	1.637** (0.744)	1.535** (0.743)
Southern Europe X *Prof*	1.349*** (0.429)	1.411*** (0.431)	1.539*** (0.430)
Further interaction terms world region X *Professional*	Yes (p> .05 for all)	Yes (p> .05 for all)	Yes (p> .05 for all)
Discipline-fixed effects	No	Yes (p> .05 for all)	Yes (p> .05 for all)
Number of observations	840	840	840

Note: The baseline group are North American environmental scientists in the *Private* identity treatment. Four pre-defined world regions, Caribbean Latin America, Central Asia, Oceania and Central America, are excluded as they feature too few observations for meaningful comparisons (2, 1, 11 and 10 observations respectively). Standard errors in parentheses. Statistical significance: *** p<0.01, ** p<0.05, * p<0.1.

[35] This positive correlation between tail reports and risk-taking is consistent with findings in Experiment I.

Result 7: Reporting behavior under professional identity priming, Experiment II

Our treatment effect in Experiment I, lower average reported tail tosses in "Professional", only replicates in Experiment II for the close baseline of environmental scientists from North America, while we cannot detect treatment effects for the majority of world regions. For Eastern Asian and Southern European scientists, we find the opposite effect.

There are several empirical investigations that report that simple experimental truth-telling tasks like the coin-tossing task we borrowed from Abeler et al. (2014) carry external validity (see Cohn et al., 2015; Cohn and Maréchal, 2018; Dai et al., 2018; Drupp et al., 2019; Gächter and Schulz, 2016; Potters and Stoop, 2016). As Experiment II includes a number of responses from several countries, we examine whether our tail-toss measure correlates with civic (dis)honesty evidence from the lost wallet field experiment by Cohn et al. (2019) and the scientific misconduct ranking in Ataie-Ashtiani (2018), both at the country-level. Figures 5 and 6 provide scatterplots and a fitted line of the data, including countries for which our dataset includes at least ten observations and which are also included in Cohn et al. (2019)'s and Ataie-Ashtiani (2018)'s dataset respectively. The correlation coefficients are -0.4953 (p=0.031) for Cohn et al. (2019) and -0.3753 (p=0.078) for Ataie-Ashtiani (2018). As returned wallets in Cohn et al. (2019) are a field measure of honesty and high tail tosses in our task are a measure of dishonesty, the results are consistent with each other. Further, we find that the higher a country's ranking is concerning scientific misconduct in Ataie-Ashtiani (2018), the higher is the average number of reported winning tails in our experiment (with China leading the ranking at place 1). We regard these findings as further evidence for external validity of coin tossing and die rolling tasks (as reviewed by Abeler et al., 2019). It suggests that a society's honesty norms may spill over and affect (dis)honest conduct of scientific research.

Result 8: Country-level reliability of the tail-toss measure of dishonesty

Our tail-toss measure of dishonesty of scientists significantly correlates with country-level measures of dishonesty, i.e. with the natural field experiment measure of civic honesty of Cohn et al. (2019) and the scientific-misconduct measure of Ataie-Ashtiani (2018).

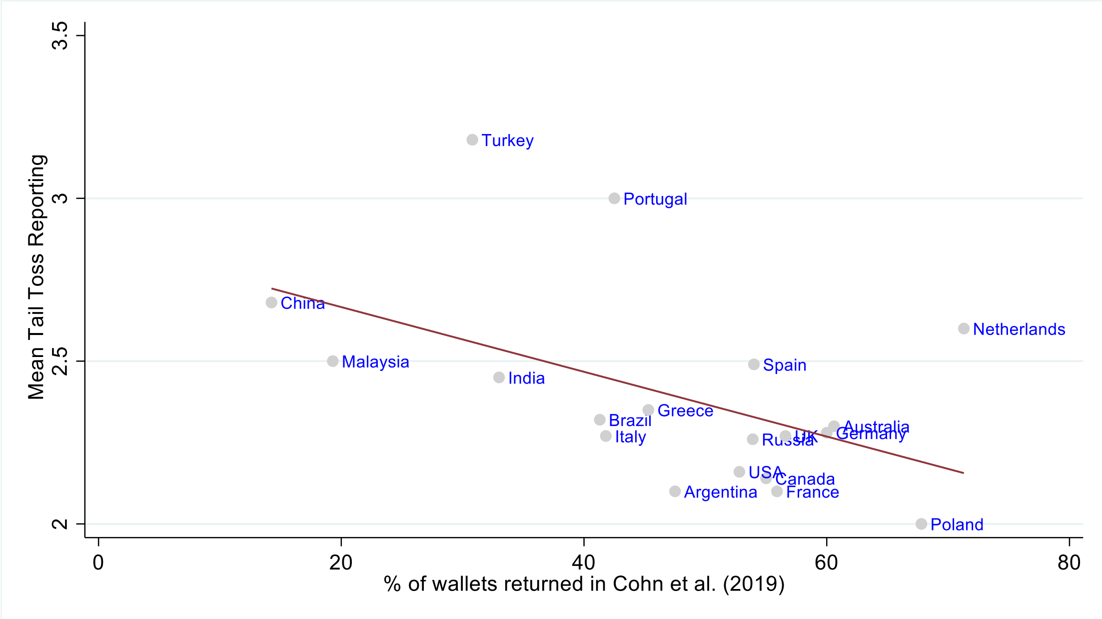


Figure 5: Correlation between Cohn et al. (2019)'s measure of (dis)honesty in the lost wallet experiment and our tail toss measure at the country-level.

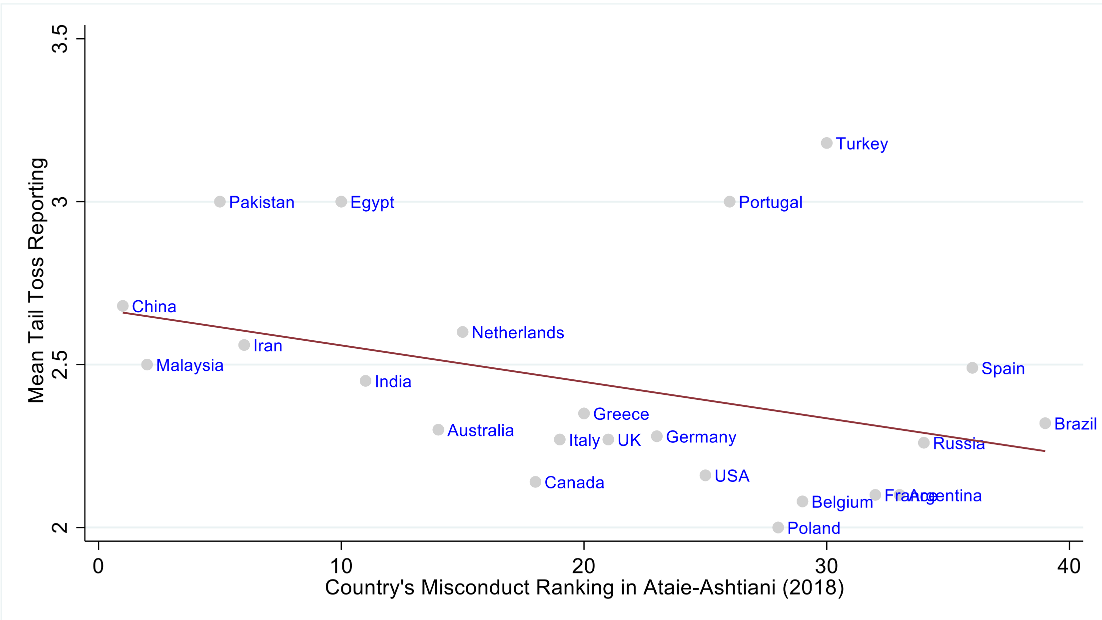


Figure 6: Correlation between Ataie-Ashtiani (2018)'s measure of scientific misconduct and our tail toss measure at the country-level.

4. Discussion and conclusion

We have investigated whether scientists, as professional seekers of truths, tell the truth by means of an incentivized coin-toss truth-telling task in two online (field) experiments with more than 1,300 scientists. In particular, we compare truth-telling behavior, in the form of coin-toss reporting, across two treatments that either made participants' professional or private identity more salient using nine identity priming questions.

Our key result in Experiment I (with marine scientists from North America and Europe) is that fewer participants over-report winning tail tosses in the professional identity treatment. In global Experiment II, we fail to uniformly replicate this result and can only confirm it for North American scientists. Overall, we find heterogeneity for the treatment effect and overreporting of scientists between world regions—reaching from no detectable difference from expected truthful reporting among Northern and Eastern European scientists to a very clear average over-reporting from scientists in some other world regions. We find a significant correlation between (dis)honesty in the general public measured by the lost-wallet field experiment by Cohn et al. (2019) and the scientists in Experiment II for a sample of 19 countries. Likewise, our reported tail tosses measure correlates with country-level measures of scientific misconduct. Our results thereby add further group-level external validity to truth-telling tasks discussed in Abeler et al. (2019).

While we provide some nuanced evidence that professional identity effects associated with science may foster truth-telling, at least in North America and parts of Europe, we can pinpoint the underlying mechanism for this finding only inductively.[36] Previous work that our simple model of truth-telling behavior builds upon (Benjamin et al., 2010; Cohn et al., 2015) suggests that this more frequent truth-telling is driven by stronger honesty norms associated with the professional (in this case scientists') identity. This main interpretation would suggest that academia is able to foster a culture of truth-telling that is consistent with its general aim of searching for truths. Indeed, this cultural norm-based interpretation has featured prominently in related findings in experimental studies on the banking industry (Cohn et al., 2014; Villeval, 2014) and it is consistent with the cross-country comparison between Cohn et al. (2019)'s results and ours. Stronger honesty norms may however not be the only facet of the professional identity of scientists that drives truth-telling behavior. For example, it is often suggested that competitiveness ("publish or perish") is a central feature of behavioral patterns and thus perhaps also

[36] Taking the study by Cohn et al. (2014) as an example, Vranka and Houdek (2015) discuss the difficulty of pinpointing underlying mechanisms of observed priming effects.

associated norms in academia (see, e.g., Fanelli, 2010; Necker, 2014). If this were the case, our main treatment effect finding would be a conservative estimate of the truth-telling norms that science nurtures, as also inherent competitiveness norms might have a detrimental effect on truth-telling.[37]

Besides the interpretation that honesty norms associated with the scientific identity affect truth-telling behavior, it could also be the case that other professional identity concerns may impact our results. Specifically, it could be that scientists strategically report more honestly as they might seek to paint a more positive picture of science. That is, they may take reputational concerns at the level of the profession into account.[38] We regard this alternative explanation as an unlikely mechanism. A necessary condition for this strategic influence explanation is that participating scientists believe that they can favorably influence the overall outcome, i.e. their contribution is non-marginal. The participants in our experiments knew that we targeted a large number of observations, i.e. 1/n was small. Given our between-subjects design, participants were also not aware that there was another treatment.[39] Thus, even though we cannot rule out the role of professional reputation concerns by design, it seems unlikely that this will be a main driver of our effects.[40]

While the indication for treatment effect of professional identity salience being associated with lower over-reporting from Experiment I seems to suggest that science may foster a culture of honesty, the heterogeneity of treatment effects and especially the over-reporting in some world regions in Experiment II seems concerning. While the professional identity priming effect in Experiment II was not comparably strong, this may suggest that existing cultures of honesty within academia are not sufficient to ensure that science does not get derailed from its quest for truths. Indeed, we overall find that scientists significantly over-report winning tail tosses on average, just as other populations (Abeler et al., 2019). This finding is in line with anonymous survey-based approaches that provide

---

[37] For example, Shleifer (2004) discusses how (market) competition may have detrimental effects on ethical behavior. More recently, a series of experimental studies have found that competition may lead to more dishonesty (see, e.g., Cartwright and Menezes, 2014; Faravelli et al., 2015; Rigdon and D'Esterre, 2015; Schwieren and Weichselbaumer, 2010). Yet, while Fanelli et al. (2015) find that scientific misconduct is more likely in countries where individual research output yields monetary rewards, their results do not support the hypothesis that pressure to publish seems to drive dishonest behavior. Furthermore, Cohn et al. (2014) do not find an identity priming effect for bankers on a stated preference question on competitiveness.

[38] This strategic behavior could thus be present in both treatments, but due to our experimentally induced higher salience it would likely be higher in the professional identity treatment.

[39] While truth-telling approaches are well-known in behavioral economics and psychology by now, the participating natural scientists in Experiment I had limited exposure to such experiments.

[40] If portraying a positive image of science would drive our treatment effect in truth-telling behavior, one might also expect such strategic behavior to show up in subsequent donation decisions. Yet, we find no significant differences across the two treatments for both the fraction of pay-off reported and for the absolute size of donations in Experiment I.

evidence that a non-negligible fraction of scientists engage in questionable research practices (see, e.g., Fanelli, 2009; John et al., 2012; List et al., 2001; Martinson et al., 2005; Necker, 2014). Relatedly, our findings on associations between over-reporting in our global Experiment II and country-level measures of academic and field-experimental dishonesty suggest that country-level norms on (dis-)honesty and associated variations in social capital (Tannenbaum et al., 2022) affect the truth-telling of scientists across regions.

As scientific honesty is crucial for scientific development as well as the public's trust in the results of science, further measures have to be taken to prevent scientific misconduct and to enhance cultures of truthful truth-seeking across scientific cultures. Meta-analyses (e.g. Abeler et al., 2019; Brodeur et al., 2016), replication studies (e.g. Camerer et al., 2016; Dreber et al., 2015; Open Science Collaboration, 2015), more precise and transparent reporting practices (e.g. Christensen and Miguel, 2018; Miguel et al., 2014; Nosek et al., 2015; Simmons et al., 2011) as well as institutional incentives and arrangement for research integrity (Butera et al., 2020; Titus et al., 2008; Titus and Bosch, 2010) are some important steps into this direction.

References

Abeler, J., Becker, A., & Falk, A. (2014). Representative evidence on lying costs. *Journal of Public Economics*, 113, 96-104.

Abeler, J., Nosenzo, D., & Raymond, C. (2019). Preferences for Truth-Telling. *Econometrica*, 87(4), 1115-1153.

Akerlof, G.A., & Kranton, R.E. (2000). Economics and identity. *Quarterly Journal of Economics*, 115(3), 715-753.

Akerlof, G.A., & Kranton, R.E. (2005). Identity and the Economics of Organizations. *Journal of Economic Perspectives*, 19(1): 9-32.

Ataie-Ashtiani, B. (2018). World map of scientific misconduct. *Science and Engineering Ethics*, 24(5), 1653-1656.

Benjamin, D., Choi, J., & Strickland, J.A. (2010). Social Identity and Preferences. *American Economic Review*, 100(4), 1913-28.

Benjamin, D.J., Choi, J.J., & Fisher, G. (2016). Religious identity and economic behavior. *Review of Economics and Statistics*, 98(4), 617-637.

Bhattacharjee, Y. (2013). The Mind of a Con Man. New York: *New York Times Magazine* article, accessed online February 10, 2017: www.nytimes.com/2013/04/28/magazine/diederik-stapels-audacious-academic-fraud.html.

Binswanger, H.P. (1981). Attitudes toward risk: Theoretical implications of an experiment in rural India. *Economic Journal*, 91, 867-890.

Blanco-Perez, C., & Brodeur, A. (2020). Publication bias and editorial statement on negative findings. *The Economic Journal*, *130*(629), 1226-1247.

Brodeur, A., Lé, M., Sangnier, M., & Zylberberg, Y. (2016). Star wars: The empirics strike back. *American Economic Journal: Applied Economics*, *8*(1), 1-32.

Brodeur, A., Cook, N., & Heyes, A. (2020). Methods matter: P-hacking and publication bias in causal analysis in economics. *American Economic Review*, *110*(11), 3634-60.

Butera, L., Grossman, P. J., Houser, D., List, J. A., & Villeval, M. C. (2020). A new mechanism to alleviate the crises of confidence in science-with an application to the public goods game. *NBER Working Paper* No. w26801.

Cadsby, C. B., Du, N., & Song, F. (2016). In-group favoritism and moral decision-making. *Journal of Economic Behavior & Organization*, 128, 59-71.

Cartwright, E., & Menezes, M. L. (2014). Cheating to win: Dishonesty and the intensity of competition. *Economics Letters, 122(1), 55*-58.

Conrads, J., Irlenbusch, B., Rilke, R. M., Schielke, A., & Walkowitz, G. (2014). Honesty in tournaments. *Economics Letters*, 123(1), 90-93.

Camerer, C.F., Dreber, A., Forsell, E., Ho, T.H., Huber, J., Johannesson, M., Kirchler, M., Almenberg, J., Altmejd, A., Chan, T. and Heikensten, E. (2016). Evaluating replicability of laboratory experiments in economics. *Science*,351(6280), 1433-1436.

Cappelen, A. W., Sørensen, E.Ø., & Tungodden, B. (2013). When do we lie? *Journal of Economic Behavior & Organization,* 93, 258-265.

Chen, Y., & Li, S. X. (2009). Group identity and social preferences. *The American Economic Review*, 99(1), 431-457.

Christensen, G.S., & Miguel, E. (2018). Transparency, Reproducibility, and the Credibility of Economics Research. *Journal of Economic Literature,*56(3), 920-80.

Cohn, A., Fehr, E., & Maréchal, M.A. (2014). Business culture and dishonesty in the banking industry. *Nature*, 516, 86–89.

Cohn, A., Fehr, E., & Maréchal, M.A. (2017): Do professional norms in the banking industry favor risk-taking? *Review of Financial Studies*, 30(11), 3801–3823.

Cohn, A., and Maréchal, M.A. (2016). Priming in Economics. *Current Opinion in Psychology*, 12, 17-21.

Cohn, A., and Maréchal, M.A. (2018). Laboratory Measure of Cheating Predicts Misbehavior at School. *Economic Journal*, 128 (615), 2743–2754.

Cohn, A., Maréchal, M.A., & Noll, T. (2015). Bad boys: How criminal identity salience affects rule violation. *Review of Economic Studies*, 82(4), 1289-1308.

Cohn, A., Maréchal, M. A., Tannenbaum, D., & Zünd, C. L. (2019). Civic honesty around the globe. *Science*, 365(6448), 70-73.

Dai, Z., Galeotti, F., and Villeval, M.C. (2018). Dishonesty in the lab predicts dishonesty in the field. An experiment in public transportations. *Management Science*, 64(3), pp. 1081–1100.

Dreber, A., Pfeiffer, T., Almenberg, J., Isaksson, S., Wilson, B., Chen, Y., Nosek, B.A., & Johannesson, M., 2015. (2015). Using prediction markets to estimate the reproducibility of scientific research. *Proceedings of the National Academy of Sciences*, 112(50), 15343-15347.

Drupp, M.A., Khadjavi, M., & Quaas, M.F. (2019). Truth-telling and the regulator. Experimental evidence from commercial fishermen. *European Economic Review*, 120, 103310.

Drupp, M.A., Khadjavi, M., Riekhof, M.-C., & Voss, R. (2020). Professional identity and the gender gap in risk-taking. Evidence from field experiments with scientists. *Journal of Economic Behavior & Organization*, 170, 418-432.

Eckel, C.C., & Grossman, P.J. (2002). Sex differences and statistical stereotyping in attitudes toward financial risk. *Evolution and human behavior* 23(4), 281-295.

Fanelli, D. (2009). How many scientists fabricate and falsify research? A systematic review and meta-analysis of survey data. *PloS one*, *4*(5), e5738.

Fanelli, D. (2010). Do pressures to publish increase scientists' bias? An empirical support from US States Data. *PloS one*, *5*(4), e10271.

Fanelli, D., Costas, R., & Larivière, V. (2015). Misconduct policies, academic culture and career stage, not gender or pressures to publish, affect scientific integrity. *PLoS One*, 10(6), e0127556.

Faravelli, M., Friesen, L., & Gangadharan, L. (2015). Selection, tournaments, and dishonesty. *Journal of Economic Behavior & Organization*, 110, 160-175.

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

Gächter, S., & Schulz, J. F. (2016). Intrinsic honesty and the prevalence of rule violations across societies. *Nature*, 531 (7595), 496-499.

Gibson, R., Tanner, C., & Wagner, A. F. (2013). Preferences for truthfulness: Heterogeneity among and within individuals. *American Economic Review*, 103, 532-548.

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

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

Gneezy, U., Kajackaite, A., & Sobel, J. (2018). Lying Aversion and the Size of the Lie. *American Economic Review*, 108(2), 419- 453.

Houser, D., Vetter, S., & Winter, J. (2012). Fairness and cheating. *European Economic Review*, 56, 1645–1655.

John, L. K., Loewenstein, G., & Prelec, D. (2012). Measuring the prevalence of questionable research practices with incentives for truth telling. *Psychological Science*, 0956797611430953.

List, J. A., Bailey, C. D., Euzent, P. J., & Martin, T. L. (2001). Academic economists behaving badly? A survey on three areas of unethical behavior. *Economic Inquiry*, 39(1), 162-170.

López-Pérez, R., & Spiegelman, E. (2013). Why do people tell the truth? Experimental evidence for pure lie aversion. *Experimental Economics*, 16(3), 233-247.

Maniadis, Z., Tufano, F., & List, J. A. (2017). To Replicate or Not to Replicate? Exploring Reproducibility in Economics Through the Lens of a Model and a Pilot Study. *The Economic Journal*, *127*(605), F209-F235.

Marshall, E. (2000). How prevalent is fraud? That's a million-dollar question. *Science*, 290(5497), 1662-1663.

Martinson, B. C., Anderson, M. S., & De Vries, R. (2005). Scientists behaving badly. *Nature*, 435(7043), 737-738.

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.

Merriam-Webster Dictionary (2017). *Science*. Accessed online on February 10, 2017: https://www.merriam-webster.com/dictionary/science.

Miguel, E., Camerer, C., Casey, K., Cohen, J., Esterling, K.M., Gerber, A., Glennerster, R., Green, D.P., Humphreys, M., Imbens, G., Laitin, D., Madon, T., Nelson, L., Nosek, B.A., Petersen, M., Sedlmayr, R., Simmons, J.P., Simonsohn, U., & Van der Laan, M. (2014). Promoting Transparency in Social Science Research. *Science*, 343(6166), 30-31.

Necker, S. (2014). Scientific Misbehavior in Economics. *Research Policy* 43, 1747-1759.

Nosek, B.A., Alter, G., Banks, G.C., Borsboom, D., Bowman, S.D., Breckler, S.J., Buck, S., Chambers, C.D., Chin, G., Christensen, G., Contestabile, M., Dafoe, A., Eich, E., Freese, J., Glennerster, R., Goroff, D., Green, D.P., Hesse, B., Humphreys, M., Ishiyama, J., Karlan, D., Kraut, A., Lupia, A., Mabry, P., Madon, T.A., Malhotra, N., Mayo-Wilson, E., McNutt, M., Miguel, E., Paluck, E.L., Simonsohn, U., Soderberg, C., Spellman, B.A., Turitto, J., VanderBos, G., Vazire, S., Wagenmakers, E.J., Wilson, R., & Yarkoni., T. (2015). Promoting an open research culture: Author guidelines for journals could help to promote transparency, openness, and reproducibility. *Science*, 348(6242), 1422-1425.

O'Grady, C. (2021). Fraudulent data raise questions about superstar honesty researcher. *Science*, 373, 950-1.

Open Science Collaboration (2015). Estimating the reproducibility of psychological science. *Science*, *349*(6251), aac4716.

Poland, G.A. & Jacobson, R.M. (2011). The Age-Old Struggle against the Antivaccinationists. *New England Journal of Medicine*, 364, 97-99.

Popper, K.R. (1996). *In search of a better world: Lectures and essays from thirty years*. Psychology Press. 245 pages.

Potters, J., & Stoop, J. (2016). Do cheaters in the lab also cheat in the field? *European Economic Review*, 87, 26-33.

Rahwan, Z., Yoeli, E., & Fasolo, B. (2019). Heterogeneity in banker culture and its influence on dishonesty. *Nature*, 575(7782), 345-349.

Rigdon, M. L., & D'Esterre, A. P. (2015). The effects of competition on the nature of cheating behavior. *Southern Economic Journal*, 81(4), 1012-1024.

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.

Sang-Hun, C. (2009). Disgraced Cloning Expert Convicted in South Korea. New York: *New York Times* article, online access on February 10, 2017: http://www.nytimes.com/2009/10/27/world/asia/27clone.html.

Schwieren, C., & Weichselbaumer, D. (2010). Does competition enhance performance or cheating? A laboratory experiment. *Journal of Economic Psychology*, 31(3), 241-253.

Shih, M., Pittinsky, T.L., & Ambady, N. (1999). Stereotype susceptibility: Identity salience and shifts in quantitative performance. *Psychological Science*, 10(1), 80-83.

Shleifer, A. (2004). Does Competition Destroy Ethical Behavior?. *The American Economic Review*, 94(2), 414-418.

Simmons, J.P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science*, 22(11), 1359-1366.

Sutter, M. (2009). Deception through telling the truth?! Experimental evidence from individuals and teams. *Economic Journal*, 119(534), 47-60.

Tannenbaum, D., Cohn, A., Zünd, C. L., & Maréchal, M. A. (2022). What do cross-country surveys tell us about social capital?. *Review of Economics and Statistics*, 1-30.

The New Yorker (2023). They Studied Dishonesty. Was Their Work a Lie? Article available online at: https://www.newyorker.com/magazine/2023/10/09/they-studied-dishonesty-was-their-work-a-lie.

Titus, S. L., Wells, J. A., & Rhoades, L. J. (2008). Repairing research integrity. *Nature*, 453(7198), 980-982.

Titus, S., & Bosch, X. (2010). Tie funding to research integrity. *Nature*, 466(7305), 436-437.

Villeval, M.C. (2014). Behavioural economics: Professional identity can increase dishonesty. *Nature*, 516(7529), 48-49.

Vranka, M.A., & Houdek, P. (2015). Many faces of bankers' identity: how (not) to study dishonesty. *Frontiers in Psychology*, 6.

Wade, N. (2010). Harvard Finds Scientist Guilty of Misconduct. New York: *New York Times* article, accessed online on February 10, 2017: http://www.nytimes.com/2010/08/21/education/21harvard.html.

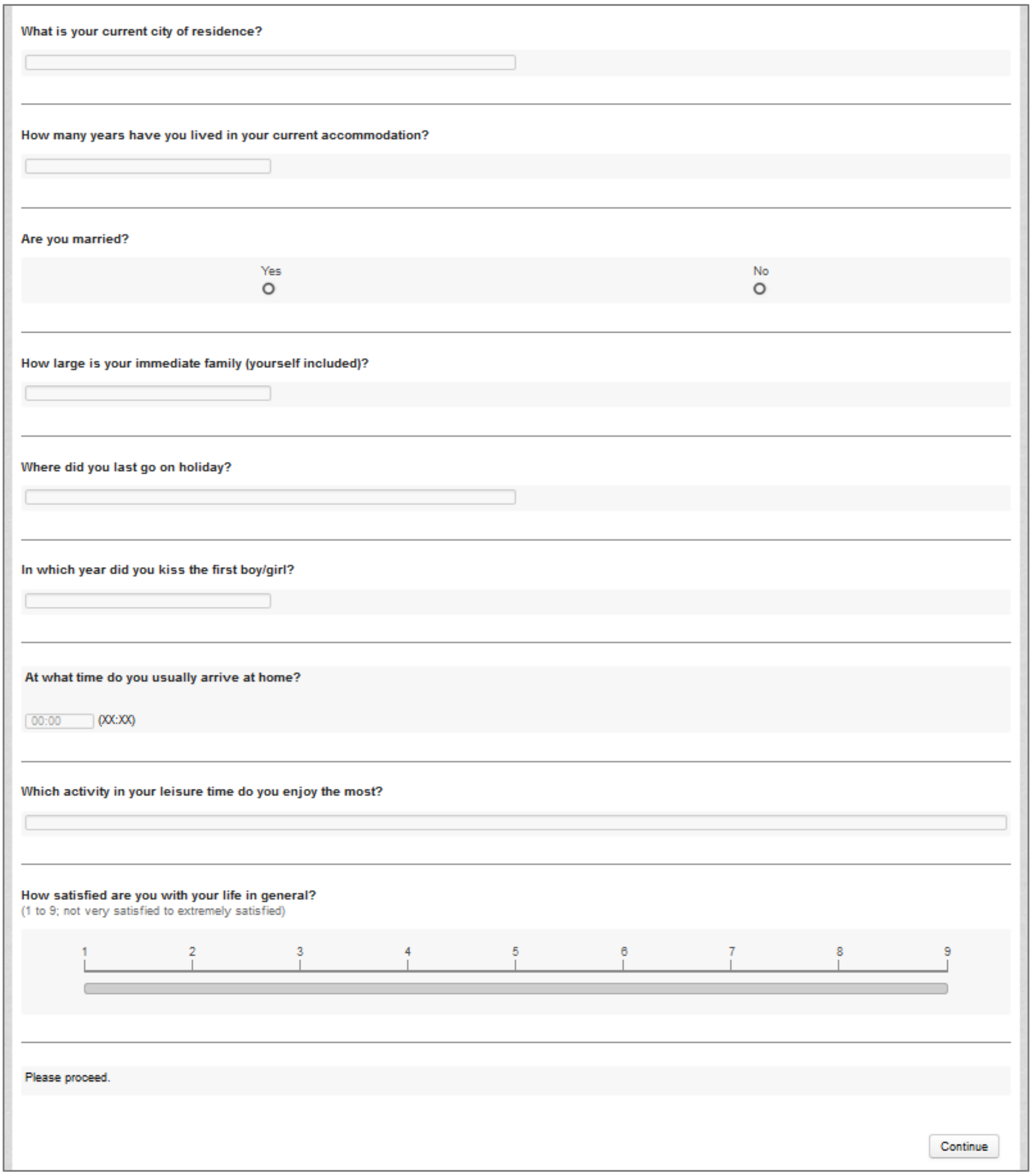
Wakefield, A.J., Murch, S.H., Anthony, A., Linnell, J., Casson, D.M., Malik, M., Berelowitz, M., Dhillon, A.P., Thomson, M.A., Harvey, P., Valentine, A., Davies, S.E., Walker-Smith, J.A. (1998). Ileal-lymphoid-nodular hyperplasia, non-specific colitis, and pervasive developmental disorder in children. *The Lancet*, 351(9103), 637-641.

Young, E. (2012). Nobel laureate challenges psychologists to clean up their act. *Nature News*.

Zimerman, L., Shalvi, S., & Bereby-Meyer, Y. (2014). Self-reported ethical risk taking tendencies predict actual dishonesty. *Judgment and Decision Making*, *9*(1), 58.

Appendix A: Screenshots from the online survey

Figure A.1: Priming questions for the *Private* treatment in Experiment I.



What is your current city of residence?

How many years have you lived in your current accommodation?

Are you married?

Yes No

How large is your immediate family (yourself included)?

Where did you last go on holiday?

In which year did you kiss the first boy/girl?

At what time do you usually arrive at home?

00:00 (XX:XX)

Which activity in your leisure time do you enjoy the most?

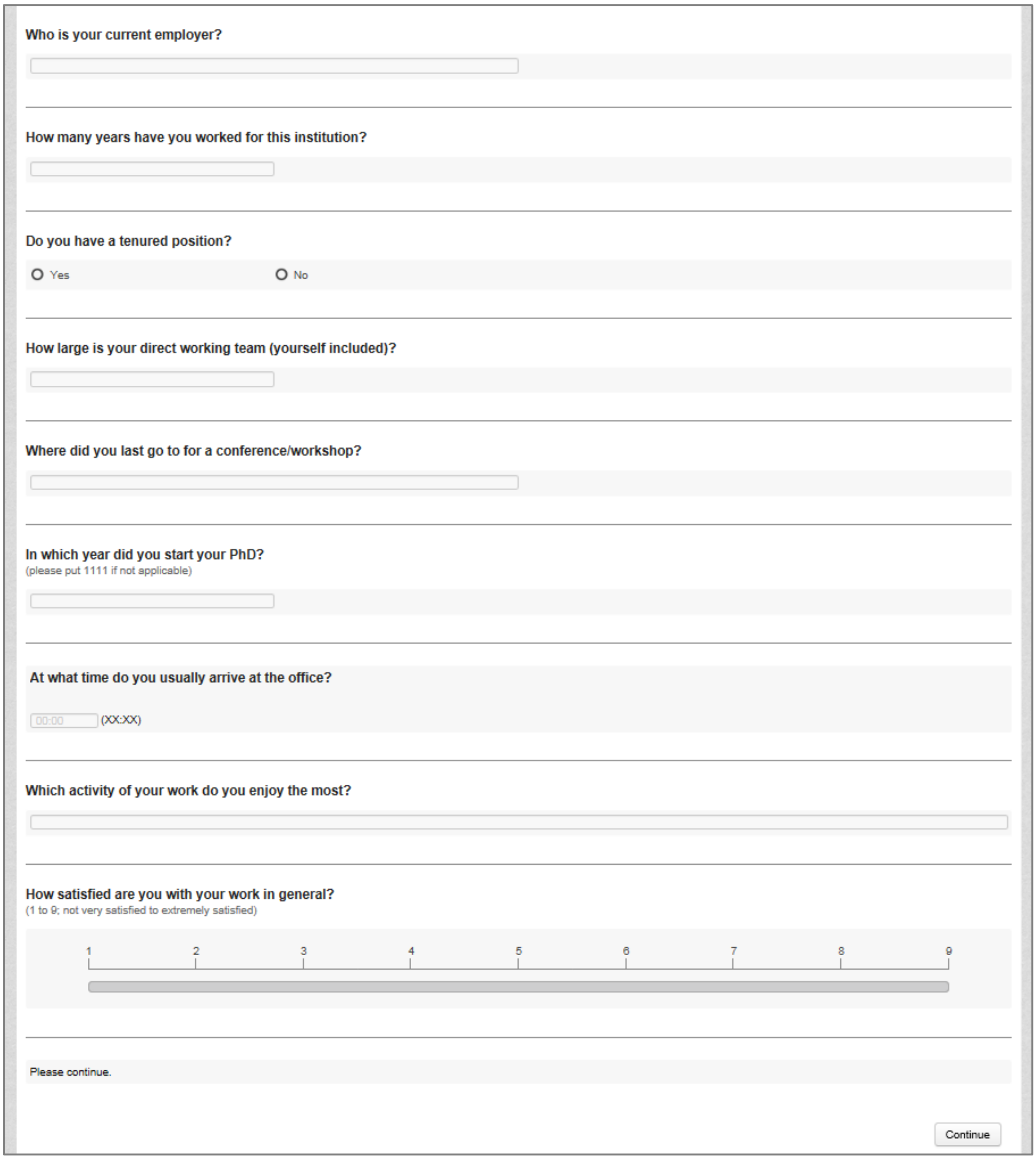
How satisfied are you with your life in general?
(1 to 9; not very satisfied to extremely satisfied)

1 2 3 4 5 6 7 8 9

Please proceed.

Continue

Figure A.2: Priming questions for the *Professional* treatment in Experiment I.



Who is your current employer?

How many years have you worked for this institution?

Do you have a tenured position?

Yes No

How large is your direct working team (yourself included)?

Where did you last go to for a conference/workshop?

In which year did you start your PhD?
(please put 1111 if not applicable)

At what time do you usually arrive at the office?

00:00 (XX:XX)

Which activity of your work do you enjoy the most?

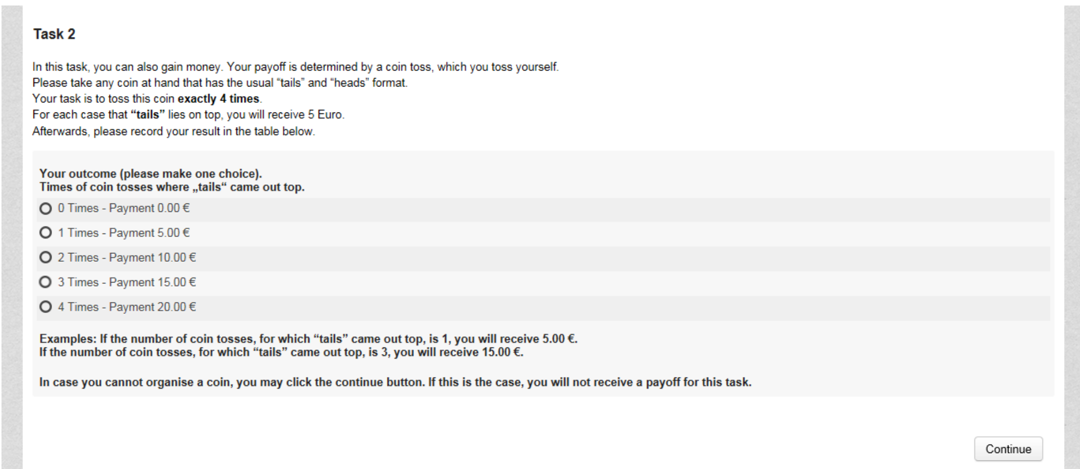
How satisfied are you with your work in general?
(1 to 9; not very satisfied to extremely satisfied)

1 2 3 4 5 6 7 8 9

Please continue.

Continue

Figure A.3: Screenshot for the coin toss-reporting task.



Task 2

In this task, you can also gain money. Your payoff is determined by a coin toss, which you toss yourself.
Please take any coin at hand that has the usual "tails" and "heads" format.
Your task is to toss this coin **exactly 4 times**.
For each case that **"tails"** lies on top, you will receive 5 Euro.
Afterwards, please record your result in the table below.

Your outcome (please make one choice).
Times of coin tosses where „tails" came out top.

- 0 Times - Payment 0.00 €
- 1 Times - Payment 5.00 €
- 2 Times - Payment 10.00 €
- 3 Times - Payment 15.00 €
- 4 Times - Payment 20.00 €

Examples: If the number of coin tosses, for which "tails" came out top, is 1, you will receive 5.00 €.
If the number of coin tosses, for which "tails" came out top, is 3, you will receive 15.00 €.

In case you cannot organise a coin, you may click the continue button. If this is the case, you will not receive a payoff for this task.

Continue

Figure A.4: Histograms of reported tail tosses by discipline in Experiment II.

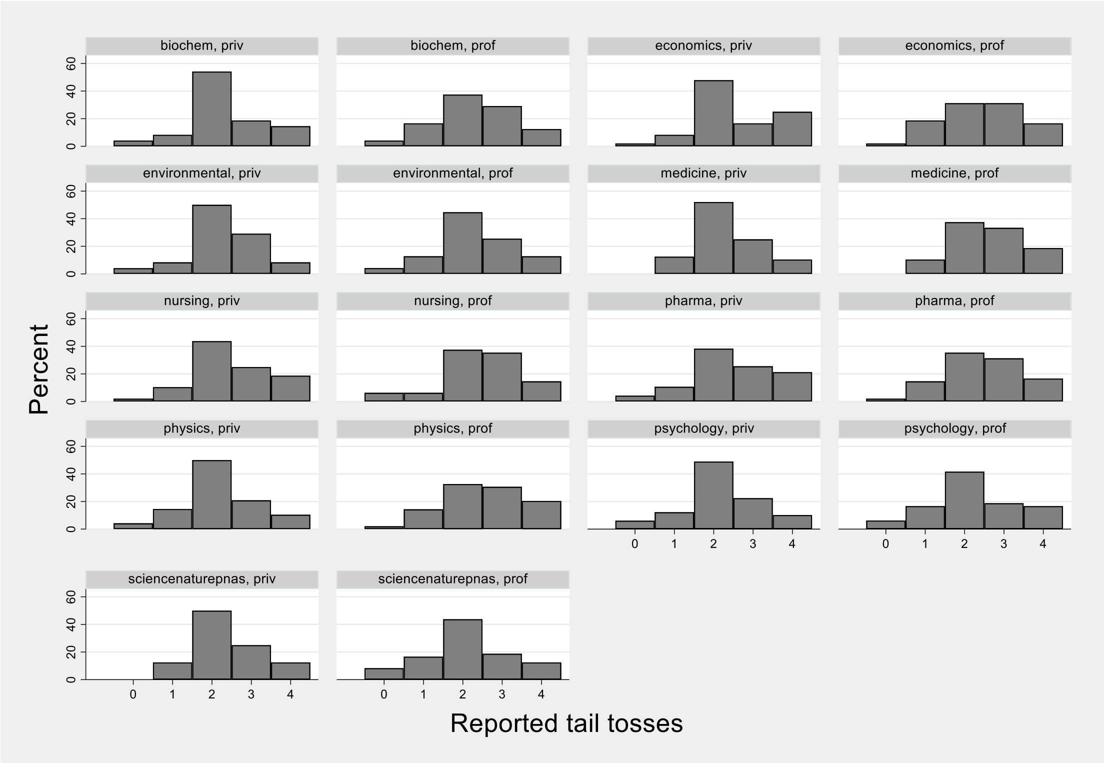


Figure A.5: Histograms of reported tail tosses by world regions in Experiment II.

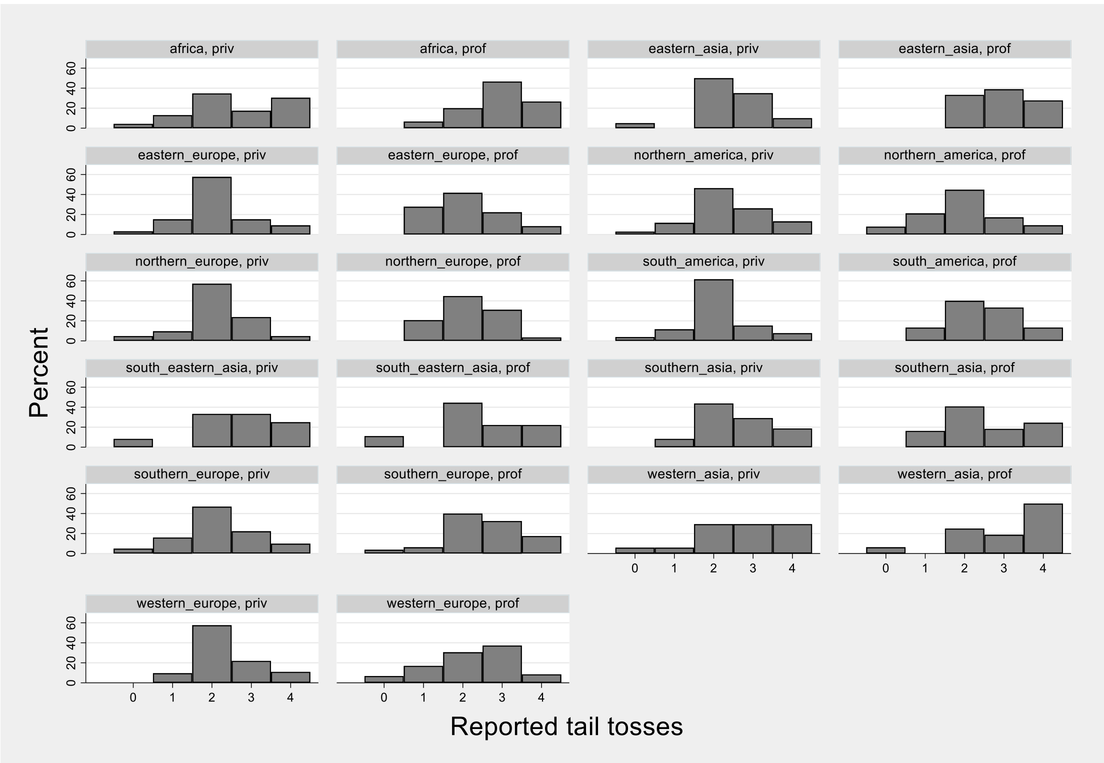


Table A.1: Priming questions for *Professional* and *Private* in Experiment II.

***Professional* Identity Treatment**	***Private* Identity Treatment**
Who is your current employer?	What is your current city of residence?
How many years have you worked for this employer?	How many years have you lived in your current accommodation?
How large is your direct working team (yourself included)?	How large is your circle of close friends (yourself included)?
Where did you last go to for a conference/workshop?	Where did you last go on holiday?
Do you coordinate your work hours with your colleagues?	Do you coordinate your work hours with your close friends?
How satisfied are you with your professional life in general? (1 to 9)	How satisfied are you with your private life in general? (1 to 9)
What part of your work do you enjoy the most? (bullet points are sufficient)	What part of your leisure time do you enjoy the most? (bullet points are sufficient)
How many hours per week do you usually spend in the office?	How many hours per week do you usually sleep?
What is your favourite academic journal?	What is your favourite newspaper?