

Schudy, Simeon; Grundmann, Susanna; Spantig, Lisa

Working Paper

Individual Preferences for Truth-Telling

CESifo Working Paper, No. 11521

Provided in Cooperation with:

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

Suggested Citation: Schudy, Simeon; Grundmann, Susanna; Spantig, Lisa (2024) : Individual Preferences for Truth-Telling, CESifo Working Paper, No. 11521, CESifo GmbH, Munich

This Version is available at:

<https://hdl.handle.net/10419/308417>

Standard-Nutzungsbedingungen:

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

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

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

Terms of use:

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

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

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

Individual Preferences for Truth-Telling

Simeon Schudy (r) Susanna Grundmann (r) Lisa Spantig

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>

Individual Preferences for Truth-Telling

Abstract

Contrary to the traditional economic view that individuals misreport private information to maximize material payoffs, recent evidence highlights robust preferences for truth-telling among many decision-makers. Theoretical models that align with aggregate behavioral patterns posit that these preferences arise from both an intrinsic motivation to be honest and a desire to be perceived as honest. We propose a novel incentivized measure to independently capture these two motives at the individual level for the first time. We validate the measure's properties experimentally and show that it predicts behavior in other commonly studied situations that allow for (dis)honesty. The measure enables the classification of individual preference types, revealing systematic heterogeneity and fairly stable type distributions across different samples. Additionally, we propose an experimentally validated 2-minute survey module that proxies both motives and predicts behavior in a typical reporting task. Including this module in a large panel, we offer first insights into how early-life experiences may shape preferences for being and being seen as honest.

JEL-Codes: C910, D010, D820, D910.

Keywords: honesty, lying costs, social image concerns, intentions, individual preferences.

Simeon Schudy
Ulm University / Germany
simeon.schudy@uni-ulm.de

Susanna Grundmann
University of Cologne / Germany
grundmann@wiso.uni-koeln.de

Lisa Spantig
RWTH Aachen University / Germany
spantig@expecon.rwth-aachen.de

November 30, 2024

We thank Carsten Otto for excellent research support and participants of the ASFEE 2023, Behavioral Brown Bag Seminar at LMU Munich, BEERS GATE Lyon, C-SEB Workshop 2022, CRC Retreat Schwanenwerder, Eastern-Arc Behavioral Conference, EEA-ESEM2024, ECONtribute SummerWorkshop 2021, Essex Behavioral and Development seminar, Essex Behavioral Mini-Conference, ESA Europe 2024, ESA Europe 2022 (early version), ESA Global Online Conference 2021, GATE Lyon seminar, GfEW Conference 2021, IMEBESS 2023, the Microeconomic retreats at Ohlstadt and Riederau, and the RWTH Aachen Econ seminar for comments. All authors acknowledge funding by the Diligentia Foundation. Susanna Grundmann acknowledges funding by the C-SEB Junior Start-Up Grant (Rd11-2020-JSUG-Grundm). Susanna Grundmann acknowledges that her contribution to this project was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy – EXC 2126/1– 390838866. Simeon Schudy gratefully acknowledges support by the German Research Foundation (DFG) through the CRC/TRR190 (Project number 280092119). The experiment was registered at aspredicted.org (70550, 82028, 129232, 131061, and 161756) and at OSF. It has been approved by the Ethics Commission of the Department of Economics, University of Munich (Project 2020-14)) and the Ethics Committee of the Social Sciences Department at the University of Cologne (210035SG, 220065SG).

1 Introduction

Fraudulent behavior is prevalent in markets, firms, and institutions, but often costly or even impossible to observe at the individual level. Examples include the overprovision of (or overcharging for) services in credence goods markets, the misreporting of income in tax declarations, as well as (mis)communication in teams, firms, or politics (see, e.g., Balafoutas et al. 2013, 2020; Bott et al. 2020; Kocher et al. 2018; Lang and Schudy 2023; Weisel and Shalvi 2015). Traditionally, economists have made the simplifying assumption that decision-makers will always misreport information they hold if such misreporting maximizes their material payoff. However, recent theoretical and empirical contributions (Abeler et al. 2019; Cohn et al. 2019; Fischbacher and Föllmi-Heusi 2013; Gneezy et al. 2018; Khalmetski and Sliwka 2019; Shalvi et al. 2012) have challenged this view and proposed that decision-makers have preferences for truth-telling which shape their (mis)reporting behavior. For example, Abeler et al. (2019) analyze aggregate-level data from 90 experimental studies documenting substantial lying aversion among participants. They highlight that two main motives are necessary to capture the behavioral regularities observed in aggregate reporting data: Preferences for being honest and preferences for being seen as honest.

Little is known about individual heterogeneity in these motives although understanding the interplay of both motives is relevant for theory and practice. On the one hand, equilibrium predictions in strategic settings may depend on preference types.¹ On the other hand, efficient institutions to reduce fraudulent behaviors may exploit knowledge about preference heterogeneity in underlying motives.² Finally, if researchers seek to predict individual behavior in other decision environments based on individual preferences for truth-telling, capturing both underlying motives for (dis)honesty independently appears crucial.³

This project provides a four-fold contribution that allows to substantially deepen the understanding of the different motivations underlying individual preferences for truth-telling. First, we develop and validate a novel, non-deceptive, and incentive-compatible experimental measure for individual preferences for truth-telling (IPT). This measure allows for the identification of preferences for being honest and/or being seen as honest, independently of each other and at the individual level. Second, we identify and document systematic heterogeneity in individual preferences for truth-telling in three samples. Third, following the pioneering work of Dohmen et al. (2011), we provide a validated 2-min survey module that allows to meaningfully proxy preferences for being honest and being seen

¹For example, Feess et al. (2023) argue that lying in groups may hinge on guilt sharing which is determined through group consensus and shared responsibility. As heterogeneity in preference types will affect the latter, measuring heterogeneity in preferences for truth-telling may be crucial to understand what shapes lying in groups and organizations.

²Kerschbamer and Sutter (2017) even argue that taking the coexistence of different preference types of agents into account is a prerequisite for the design of institutions and Geraedts et al. (2022) suggest that individual heterogeneity in moral capacities (rather than situational factors) may be an important driver for dishonesty.

³For instance, recent work by Grosch and Rau (2017) relates (dis)honest behavior in the die-rolling paradigm with individuals social value orientation (SVO, Murphy et al. 2011).

as honest, which can be used in large-scale population studies in which an incentivized preference elicitation is not feasible. Finally, we provide initial examples of potential research questions that can be explored with the novel survey module. To do so, we included the module in a large panel survey (SOEP-IS) and investigate how the preferences for being honest and being seen as honest relate to labor market outcomes and household formation. Using data on early life experiences, we further provide first insights into what may shape individual preferences for truth-telling.

Based on recent models of lying costs in which larger costs indicate a stronger preference for truth-telling (Abeler et al. 2019; Gneezy et al. 2018; Khalmetski and Sliwka 2019), our empirical approach conceptualizes the preference to be honest with intrinsic lying costs (ILC) that may arise due to moral or identity concerns (see, e.g., Akerlof and Kranton 2000). Further, it assumes that individuals may incur social image costs (SIC) reflecting their preferences for being seen as honest. A key empirical challenge is to measure these costs independently at the individual level. For instance, the classical die-rolling paradigm (Fischbacher and Föllmi-Heusi 2013) and other unobserved reporting tasks identify misreporting only at the aggregate level (by comparing the empirical distribution of reports to the expected distribution under truth-telling). These paradigms encompass several important features that may help to cleanly capture participants' general preferences for truth-telling⁴, but they neither allow to identify preferences for truth-telling nor the two underlying motives at the individual level. While researchers may introduce observability to measure individual dishonesty, this alters lying costs due to image concerns (see for example, Gneezy et al. 2018). As such, observability is not a useful approach to identifying individual preferences for being and being seen as honest.⁵

The novelty of our approach consists of relying on intentions instead of actual reports when measuring underlying motives for truth-telling and builds on recent work that investigates selection into different environments that do or do not allow for dishonest behavior.⁶ Thereby, we can keep all desirable features of the classical die-roll paradigm (i.e., no strategic interactions, no effort provision, and no observability of the true state by the experimenter). To measure intrinsic lying costs (ILC) and social image costs (SIC) independently of each other and in one coherent setting, we develop a novel experimental paradigm. In this paradigm, decision-makers reveal i) their intention to be dishonest (by acquiring costly information that is only useful when they plan to misreport) and ii) their intention to be seen as honest (by changing at some cost which information some independent observers who judge the decision-makers' character receive). Crucially, the experimental elicitation procedure is designed such that individuals with social image concerns can avoid the latter independently of

⁴For example, these unobserved reporting tasks avoid biases due to strategic interactions, social preferences, performance-related image concerns, or feelings of entitlement prevalent in other tasks.

⁵An alternative, but deceptive, approach is to avoid such bias by misrepresenting observability and make participants believe dishonest acts cannot be identified by the researcher (see, e.g., Albertazzi 2021; Mazar et al. 2008). However, this approach does not appear appealing when developing a measure that can be repeatedly used.

⁶See for example Fehrler et al. (2020a), Konrad et al. (2014), Lefebvre et al. (2015), and Saccardo and Serra-Garcia (2023).

whether or not they plan to misreport. Thus, the procedure can capture both underlying motives independently.⁷ We run different experimental treatments that vary exogenously whether ILC or SIC can play a role and find that participants systematically adjust their choices in our preference elicitation procedure. As such, we show that our measures of ILC and SIC have high internal validity.

Our novel paradigm elicits IPT using two incentivized willingness-to-pay measures that reflect a decision-maker's intrinsic lying costs (ILC) and social image costs (SIC). As such, the IPT measure allows for a detailed comparison of preferences across individuals and a classification of preference types. For simplicity, and following our pre-analyses plan, this paper mainly focuses on four different preference types which are defined relative to each other. The first type has low ILC (i.e., is willing to acquire information that is useful when misreporting) and low SIC (i.e., is not willing to incur costs to improve how others may judge their character). These individuals have relatively weak preferences for truth-telling. The second type has high ILC and low SIC. Hence, the second type has an intrinsic preference for truth-telling but does not care much about how others judge their behavior. The third type has low ILC, but high SIC. This type acts more honestly if they are observed by others but does not care intrinsically about being honest. The fourth type has both high ILC and high SIC and thus has relatively strong preferences for truth-telling.

Using our novel incentivized measure for IPT, we classify individual preference types in three different samples: a student sample ($n=331$), a convenience sample ($n=471$), and a representative sample ($n=500$). We document systematic heterogeneity in preferences for truth-telling. In all three samples, we find that all four above-presented preference types exist. In addition, the type distribution is fairly robust across samples. The most prevalent type has low ILC and high SIC (between 35% and 39% of the samples) while the least prevalent type has high ILC and high SIC (7% – 11%). We also observe a substantial fraction of types with low ILC and low SIC in all samples (26% – 31%) as well as types with high ILC and low SIC (21–31%).⁸ This heterogeneity underlines the importance of measuring both motivations at the individual level.

In a next step, we show that these preference types predict behavior in two other incentivized experimental paradigms, in which participants can lie to increase their payoff. These additional paradigms vary in aspects of truth-telling that have been widely studied, in particular, the observability of the true state by the experimenter and the existence of negative externalities of dishonest choices. We find that types with low ILC claim higher payoffs than types with high ILC in a misreporting paradigm without externalities (a mind game version of the die-rolling paradigm). Further, and focusing on a game with externalities, we document that a within-person variation of observability in a sender-receiver game (i.e., whether the receiver learns that she has been deceived or not) has

⁷We also test independence explicitly with an experimental treatment variation, see Section 4.1 for details.

⁸The latter type mirrors a deontological interpretation of the moral costs of lying, emphasizing principles independent of potential consequences (see, e.g., Feess et al. 2022).

stronger effects for types with high as compared to low SIC. Hence, we show that our IPT measure meaningfully relates to behavior in other situations where individuals trade off honesty and material gains. We also explore whether our IPT measure relates to behavior in a novel paradigm, in which people lie for their social image (by appearing more knowledgeable) instead of lying for monetary payoffs. We find that participants classified as having social image concerns in the knowledge reporting task are about ten percentage points more likely to be classified as high SIC individuals according to the IPT measure.⁹ Further, we find that the IPT measure yields robust results, independent of whether it is administered before or after other experimental tasks.

To quickly proxy types, we also develop a vignette-based survey measure of IPT that asks respondents to answer two simple questions that can be completed within less than 2 minutes. The answers allow to classify respondents in a binary way for both the preference for being honest and for being seen as honest. We find that the two willingness-to-pay measures and the incentivized type classification relate systematically to the response to these questions. In addition, responses to the survey questions predict behavior in a widely studied experimental misreporting paradigm. The survey measure can usefully proxy both underlying motives for individual preferences for truth-telling in large-scale studies that do not allow for the relatively more time-intensive incentivized measure.

We showcase how our survey module can be used to study different research questions by introducing it in the German socio-economic panel (SOEP-IS 2023), a large-scale representative panel survey. Exploiting the panel nature of the SOEP, we study the role of IPT in two economically relevant domains (labor market outcomes and household formation) and investigate the role of early life experiences for individual preferences for truth-telling. In terms of labor market outcomes, we find that individuals with high intrinsic lying costs are less likely to change their jobs. In terms of household formation, we document assortative matching based on IPT types. Finally, we find suggestive evidence that early life experiences such as exposure to religious parents or exposure to schooling the German Democratic Republic systematically related to individual's social image costs.

Our novel approach contributes to ongoing research on preferences for truth-telling, a dynamic and important field in economics, psychology, and the social sciences. From an experimental economics perspective, so far three seminal approaches have been used to measure the extent of (dis)honesty: i) the sender-receiver deception game (Gneezy 2005), in which informed participants can deceive uninformed participants to increase their own payoff in a strategic interaction, ii) the matrix task (see Grolleau et al. 2016; Mazar et al. 2008; Verschuere et al. 2018), in which participants can misreport their performance to increase profits at the cost of the experimenter, and, iii) the

⁹Note that our experimental measurement relies on interpreting behavioral types relative to each other. Akin to other paradigms (see e.g., Kajackaite and Gneezy 2017), behavior in the IPT paradigm may vary with stakes. Nevertheless, we observe that not only types but also the willingness-to-pay measures meaningfully predict behavior in the mind game, the sender receiver game, and the knowledge reporting task.

die-rolling paradigm (Fischbacher and Föllmi-Heusi 2013; Shalvi et al. 2012), in which participants generate a random outcome that is unobserved by the experimenter and have to report this outcome to the experimenter. All these tasks have in common that participants can benefit monetarily from misreporting and all of them show that many individuals are willing to lie but often do not lie to the full extent. The aggregate empirical regularities are best captured by, both, preferences for being and being seen as honest (Abeler et al. 2019; Gneezy et al. 2018; Khalmetski and Sliwka 2019) but previous work does not allow for studying the co-existence of different preference types. While recent valuable contributions have started to investigate the relevance of both motives focusing on how exogenously manipulating the potential role of ILC and SIC between-subjects alters behavior in the aggregate (see, e.g., Bašić and Quercia 2022), our approach allows researchers to understand the relative importance of these motivations at the individual level. We provide robust evidence on substantial heterogeneity in preference types (based on both motives) which can meaningfully enrich theoretical models as well as improve the efficacy of institutions aiming at the reduction of fraudulent behaviors. Most importantly, our results underline that people who have strong intrinsic preferences for being honest care less about how they are perceived by others than those with weak intrinsic preferences which may allow for meaningful adjustments of models that include some or both types of lying costs (see e.g. Abeler et al. 2019; M. Dufwenberg and M. A. Dufwenberg 2018; Gneezy et al. 2018; Khalmetski and Sliwka 2019). Finally, our incentivized measure will allow for direct empirical tests of theories that make type-based predictions and can be used to predict behaviors in other domains.

The remainder of this paper is structured as follows. In Section 2, we introduce our novel incentivized measure for IPT. In Section 3, we first characterize individual preferences for truth-telling based on our measures of ILC and SIC. In Section 4, we discuss the internal validity, the predictiveness, and the robustness of our measure. In Section 5, we introduce our survey measure. We show that it is systematically related to the incentivized IPT measure and find that it can also be used to meaningfully predict behavior in a mind game version of the die-rolling paradigm (Fischbacher and Föllmi-Heusi 2013). Further, we illustrate possible applications of the survey measure by including the survey module in the Innovation Sample of the German Socioeconomic Panel (SOEP-IS). Finally, in Section 6, we replicate our core findings in a representative UK sample and compares type distributions and correlates of IPT across our three different samples. Section 7 concludes.

2 Experimental Measure for IPT

2.1 Experimental Design and Assumptions

Our IPT measure builds on a reporting paradigm akin to the idea in the seminal work of Fischbacher and Föllmi-Heusi (2013). In the classical version of this paradigm, participants generate a random

outcome that is unknown to the experimenter and report it. The mapping of the report to the monetary payoffs is known before reporting. As such, participants may increase their payoff by dishonestly reporting an outcome that is associated with a higher payoff than the true outcome. As the experimenter does not know the true state, they can only infer the extent of misreporting by comparing the distribution of reports to the expected distribution. Inferring whether a given individual is honest is hence impossible with certainty and can only be approximated with a sufficiently large number of reports per individual.

Our experiment focuses on two decisions made by participants (called Decision-Makers, in short: DMs). These two decisions allow us to capture the preferences to be honest and to be seen as honest in one coherent setting but independently of each other. To capture the preference for being seen as honest, the experiment also involves Observers, who reflect on how (dis)honest DMs' behavior appears. In contrast to previous reporting paradigms, the key idea of our experiment is to measure DMs' intention to be dishonest and the intention to be seen as honest, rather than relying on the outcome the participant eventually reports. This allows us to classify DMs as having stronger or weaker preferences for being honest and appearing honest without us knowing the true state of the world. Following the experimental protocol, we first explain the general setup, and then describe how we elicit DMs' intentions.

Setup

First, DMs generate two random outcomes using an external, existing 'random pick' website. In each of the two random picks, a random device picks (with equal probability) one out of eight possible items shown to DMs. These items are 'categorical' to avoid any inherent ordering of items which might change the perceived size of the lie.¹⁰ Importantly, in our paradigm, the experimenter knows the outcome space, but does not know which items have been generated. For each of the two random picks, we instruct DMs to privately write down the outcome as they are required to report the randomly selected items at a later stage.¹¹

After participants confirmed that they generated the outcomes, we inform DMs that the two items they will report later will determine their payoffs. Decision makers are informed that different items yield different payoffs according to a pre-specified and randomly assigned payoff scheme. For one random pick, seven out of eight items yield 2 experimental currency units (ECU), while one out

¹⁰In classical reporting paradigms, reporting a '2' on a six-sided die that shows a '1' might be perceived as a smaller lie than reporting a '5' even if the payoffs these reports yield are the same (see Gneezy et al. 2018, for a demonstration of how variation in the size of the lie can impact behavior). To avoid such effects, we used fruits and vegetables as categorical items. A wheel of fortune turned by participants determined the random picks and the order of the item type (fruits or vegetables) as well as the placement of the items on the wheel of fortune were randomized at the individual level. Screenshots of the 'random pick' websites can be found in Appendix E.1.

¹¹Documenting the outcomes ensures that they are perceived as 'realized,' reducing the scope for self-deception and minimizing the potential influence of memory.

of the eight items yields 10 ECU. This is the ‘low-stakes task’. For the other random pick, seven items yield a payoff of 200 ECU, while one item yields a payoff of 1000 ECU. This is the ‘high-stakes task’. Note that in both tasks, seven items yield a lower payoff (likely, low-paying outcome) and one item yields a higher payoff (unlikely, high-paying outcome). While DMs do not yet know which items yield which payoff, they learn about the size of payoffs in the two reporting tasks. DMs are further informed that they will learn the correspondence of payoffs and reported items for at least one of the reporting tasks before reporting the item.

Finally, DMs learn that other participants are recruited as independent Observers to reflect on the DMs’ character depending on the outcome of *one* of their reports, i.e., whether their report yields the unlikely and high-paying or a more likely and low-paying outcome. Which report (high or low stake) is shown to Observers is by default randomly determined, and DMs are aware of this. To avoid curiosity confounds, DMs do not receive feedback regarding Observers’ perception of their character, and DMs know this. As such, our setup isolates pure, non-instrumental social image concerns (relating to social image costs caused by how others may perceive the DMs’ behavior).

The Observers’ task is to reflect on DMs’ character by filling in a scorecard on which they rate the probability of the DM being an honest person, whether they would trust the DM, whether they would lend the DM money and whether they would buy a used car from the DM. Observers know that they make their assessment based on reporting behavior in one reporting task, in which DMs randomly draw an item and report it. They are not informed that DMs made reports in two tasks and thus also not how the report for the assessment was selected. Observers know that the report could result in either a high or a low payment. They further know the corresponding probabilities of the likely, low-paying outcome (7/8) and the unlikely, high-paying outcome (1/8), and they are aware that DMs may have known the relationship between the report and the outcome.¹² All of the above is known to DMs. As Observers are only present to induce social image costs in DMs, we focus exclusively on DMs in the following and provide more information on Observers in Appendix A.4.

Intention to be seen as honest

DMs know that Observers are unaware that there are two reporting tasks and that the Observers will only learn whether DMs achieved the unlikely high-paying (or a likely low-paying) outcome for one of the reporting tasks. By default, DMs do not know whether Observers reflect on their behavior in the high- or low-stakes task. Previous research has shown that, in reporting tasks, less likely events that yield higher payoffs are perceived as more dishonest (see, e.g., Bašić and Quercia 2022). However, it is unclear to what extent beliefs about ratings may vary across participants. To fix DMs’

¹²To ease procedures, we use the strategy method to elicit observers’ ratings for both potential reports and provide them with information about DMs’ individual reports afterwards (see Appendix E.1).

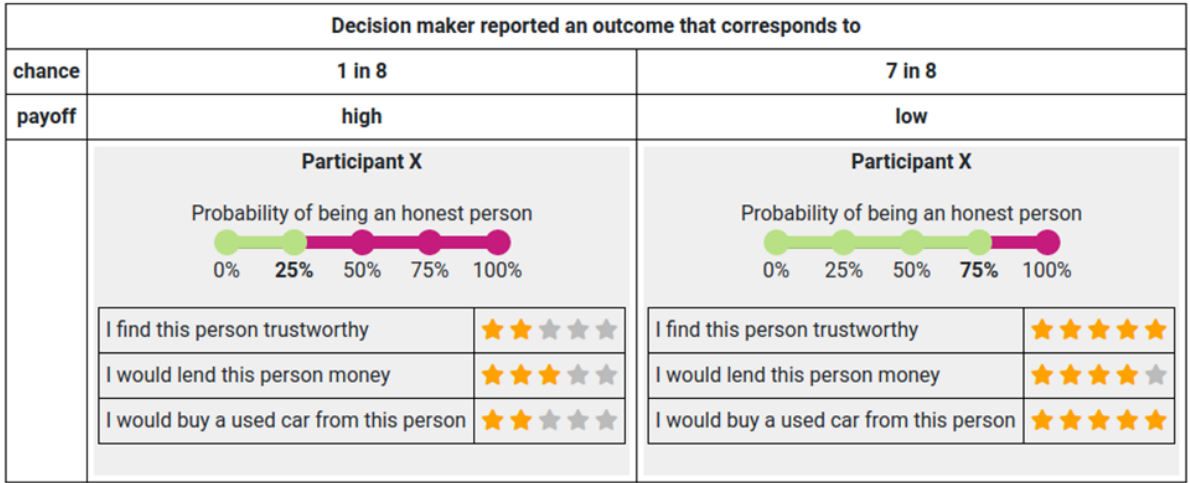


Figure 1: Scorecard shown to decision-makers

Notes: Ratings were obtained in a pilot experiment with 27 observers.

beliefs about Observers' ratings, we show DMs how Observers in the past rated DMs depending on whether their report yielded an unlikely high payoff or a more likely low payoff. These ratings are based on pre-registered pilot experiments (see Figure 1 for how this was displayed).

To measure DMs' intention to be seen as honest, we offer DMs the chance of a 'favorable switch' before making their reports. This switch implies changing the randomly selected default task to the alternative reporting task if this is expected to improve the observers' perception of the decision-makers' character. We decided to focus on a situation in which upholding a positive social image by reporting a low payoff would be very costly and thus implemented the switch only if the following conditions applied (and DMs knew this): i) the high-stakes task is selected as the default task, ii) the report in the default task coincides with the unlikely high payoff, and iii) the report in the alternative task coincides with the more likely low payoff. It becomes clear that if these conditions apply, Observers may perceive even honest DMs who were lucky in one of the two reporting tasks as dishonest and thus DMs may suffer from social image costs independently of whether they actually report (dis)honestly. We elicit DMs' willingness to pay (WTP) to switch the default task (WTP_{switch}) following G. M. Becker et al. (1964) which allows DMs to incur costs to increase the likelihood of being seen as more honest without changing or misrepresenting their actual reports. Importantly, DMs only pay for the switch if all the above-named conditions apply and thus the switch allows DMs to uphold a positive social image independent of their intrinsic lying costs.¹³

¹³This elicitation procedure minimizes the probability that the switch itself is considered immoral. First, note that the switch is favorable for any DM who cares about their social image; independent of whether the situation described above occurred due to honest or dishonest reporting. Second, the default task was randomly chosen, such that from a normative perspective, there is little reason to believe that Os are supposed to observe one particular reporting task.

Intention to be (dis)honest

After eliciting the intention to be seen as honest and before providing feedback regarding the price to be paid for the switch, we inform DMs about the correspondence of items and payoffs for the low-stakes task. That is, DMs learn which of the eight items coincides with which payoff (2 or 10 ECU) such that they are able to intentionally misreport the item the random pick website selected to increase their payoff. For the high-stakes task, DMs do not learn the correspondence of items to payoffs by default. DMs only know that one out of the eight items will be randomly assigned to the high payoff (1000 ECU) while the other seven will be assigned a low payoff (200 ECU). To measure the intention to be (dis)honest, we then elicit the WTP to learn which of the eight items will yield the high payoff (WTP_{info}) before making their report. If decision-makers do not pay for learning the information before their report, they learn the correspondence of items to payoffs in the high-stakes task directly after making their report. Hence, the WTP_{info} indicates DMs' intention to misreport. Knowing the correspondence of items to payoffs for the high-stakes task enables DMs to misreport if they are willing to do so while, by default, DMs are not able to intentionally misreport in the high-stakes task. In choosing their WTP_{info} , DMs trade off their intrinsic lying costs and the personal gain from lying. If the former exceeds the latter, WTP_{info} will be zero. In other words, if a DM intends to be honest, it is rational to not pay anything for learning the correspondence of items to payoffs before reporting. Conversely, a positive WTP_{info} indicates lower intrinsic lying costs.¹⁴

These findings underline that the core component of the WTP_{info} indeed captures variation in individual preferences for being honest (for a more detailed discussion, see also Appendix A.3). The higher WTP_{info} , the lower the intrinsic lying costs, given that the gain from lying must compensate the amount paid for the information as well as the intrinsic lying costs incurred. Importantly, as DMs were able to avoid potential social image costs (see WTP_{switch} elicitation above), the WTP_{info} measures intrinsic lying costs independently of image concerns.

Our measure for preferences for truth-telling thus captures the intention to be dishonest and to be seen as honest at the individual level. A higher WTP_{info} implies a weaker preference for being honest (i.e., lower intrinsic lying costs), whereas a higher WTP_{switch} implies a stronger preference for being seen as honest (i.e., higher social image costs).

¹⁴Note that additional analyses show that neither potential curiosity nor risk preferences affect the predictiveness of the WTP_{info} for behavior in another commonly used task to elicit honesty preferences (the mind game version of the die-rolling paradigm). Further, we designed the WTP_{info} elicitation to render the experimenter non-salient, ensuring that participants would not be concerned about the impression they make on the researcher when stating a positive willingness to pay. We also find supportive evidence for this presumption, as participants with a positive WTP_{info} report caring more about the impression they make on the researcher than those who are not willing to pay for the information. Hence, concerns regarding the experimenter do not prevent participants from seeking information (for further details, see Appendix C.2).

Report

Finally, DMs are asked to report the two items selected by the random pick website. Before reporting, DMs learn the randomly chosen default reporting task and whether their maximum WTP_{switch} and WTP_{info} were high enough to result in changes from the defaults. More specifically, if DMs' WTP_{switch} was high enough and the relevant scenario arises (see subsection 'Intention to be seen as honest'), DMs learn that the scorecard will be based on a report that results in a likely low-paying outcome in the low-paying task (if such a report exists). If their WTP_{info} was large enough, DMs learn their which item yields the high payoff in the high-stakes task before reporting. Otherwise, DMs learn the correspondence of items to payoffs for the high-stakes report directly after entering their reports.

To make a report, DMs select an item out of a randomly ordered list of the eight items.¹⁵ DMs first select their low-stakes item, and, on the same page, select their high-stakes item.

After reporting, DMs learn their payoffs from the task and Observers learn their payoffs from the rating, the scorecards their matched Observers designed, as well as for which DMs (based on the participant number) which scorecard was relevant.

2.2 Summary and procedures

To summarize, our experimental paradigm elicits DMs' individual preferences for truth-telling (IPT) based on their intention to be honest (captured by the WTP_{info}) and their intention to be seen as honest (captured by the WTP_{switch}). This allows us to provide a detailed characterization of IPT, to study heterogeneity in IPT based on both dimensions, and to classify preference types.

We elicit IPT, in a general population sample as well as in a student sample. Thereby, we are able to study the presence of heterogeneity in IPT across different samples as well as to study the internal validity and the predictability of our IPT measure (through additional experimental paradigms and treatment variations). The experiment was administered online in both samples, and programmed in oTree (Chen et al. 2016). Data for the general population sample was collected using Prolific.¹⁶

Our empirical approach encompasses several samples. We first present results from the elicitation of our incentivized measure with 471 DMs and 12 Observers from a convenience sample (general

¹⁵To avoid potential concerns of observability of the true state, pictures of items to be selected for reporting were similar but not identical to those shown on the random pick website.

¹⁶We pre-screened participants using the following criteria: English as a first language; Prolific approval rating between 90 and 100 (max); between 5 and 10000 previous submissions on Prolific. Note that we administered two treatment variations in this data collection. Both these variations focus on (presumably) changing the intensive margins of lying costs. In the first variation, we explored whether allowing DMs to spin the wheel of fortune more often while still requiring DMs to report a specific outcome (*FewSpins* vs. *ManySpins*) decreases intrinsic lying costs (akin to the approach employed by (Shalvi et al. 2011) in the die-rolling paradigm). In the second, we increased the number of observers from two to ten (*FewObservers* vs. *ManyObservers*). Both these variations were ineffective. Comparing *FewSpins* vs. *ManySpins*, we found no significant differences in WTP_{info} (MWU, $p=0.578$) nor in the share of participants with a positive WTP_{info} (χ^2 -test, $p=0.868$). Further, neither the difference in WTP_{switch} between *FewObservers* and *ManyObservers* was significant (MWU, $p=0.605$) nor the difference in the share of participants with a positive WTP_{switch} (χ^2 -test, $p=0.340$). As these treatments were part of the data collection on Prolific described in Section 3, we pooled these treatments for the main analyses.

population UK) recruited via Prolific and validate the measure’s properties in the same participant pool.¹⁷ Specifically, we recruited 499 additional Prolific participants to validate that the WTP_{info} captures DMs’ intrinsic lying costs (ILC) and another 503 Prolific participants to validate that DMs incur social image costs when having observers as compared to not having any (see Section 4). To study how well our IPT measure predicts behavior in other experimental paradigms as well as to study the robustness with respect to the timing of our elicitation procedure within a given experimental session (i.e., before or after other experimental paradigms), we also collected data from a student sample that encompasses 331 DMs and 24 Os (all participants of the Munich Experimental Laboratory for Economic and Social Sciences, MELESSA).¹⁸ Finally, we elicited the incentivized measure in a representative sample of UK Prolific participants ($n=500$, see Section 6).

3 Results

This section provides an overview of the main outcomes of the experimental measure for the general population sample. All analyses are based on DMs’ WTP_{info} and WTP_{switch} .¹⁹

3.1 Individual preferences for truth-telling: ILC and SIC

Figure 2 presents DMs’ choices for our two main outcome variables, intrinsic lying costs and social image costs.²⁰ Panel 2a provides evidence for the existence of intrinsic lying costs: 38.6% of participants are not willing to pay any positive amount to receive information on the correspondence of items and payoffs before reporting. Hence, for these participants, we do not observe an intention to lie. In contrast, 61.4% of participants have a positive willingness to pay and thus reveal an intention to lie. We observe substantial variation in the elicited amounts which illustrates strong heterogeneity in intrinsic lying costs across individuals.

Panel 2b presents evidence for social image costs. We observe that 58.0% of DMs are not willing to pay any positive amount for the favorable switch, and are thus classified as exhibiting low SIC. For the remaining 52.0%, we find variation in the elicited amounts, implying heterogeneity in the extent of SIC. Interestingly, we find that the lower a DM’s intrinsic lying costs (ILC), the higher is their SIC (Spearman’s $\rho = -0.39$, $p\text{-value} < 0.01$). While many theoretical models assume both, intrinsic lying

¹⁷Participants completed the core study (IPT measure, mind game, and questionnaire) on average in 22 minutes and their average payments amounted to £5.69.

¹⁸The main reason to recruit students in addition was the longer expected duration due to the additional paradigms. Eventually, the student sessions lasted on average 40 minutes (and average payments amounted to 17.75€, including the show-up fee of 6€).

¹⁹Results regarding IPT in the student sample are presented in Appendix Section A.1. Appendix B.2 presents results for DMs’ reports and Appendix A.4 presents observers’ behavior.

²⁰Note that participants could indicate their WTP by choosing one of the following options: 0, 25, 50, 75, 100, 125, 150, 175, 200, 250, 300, 350, 400, 500, 600, 700, 800. We elicit WTP in a discrete way to ease understanding. See also the instructions in Appendix E.1.

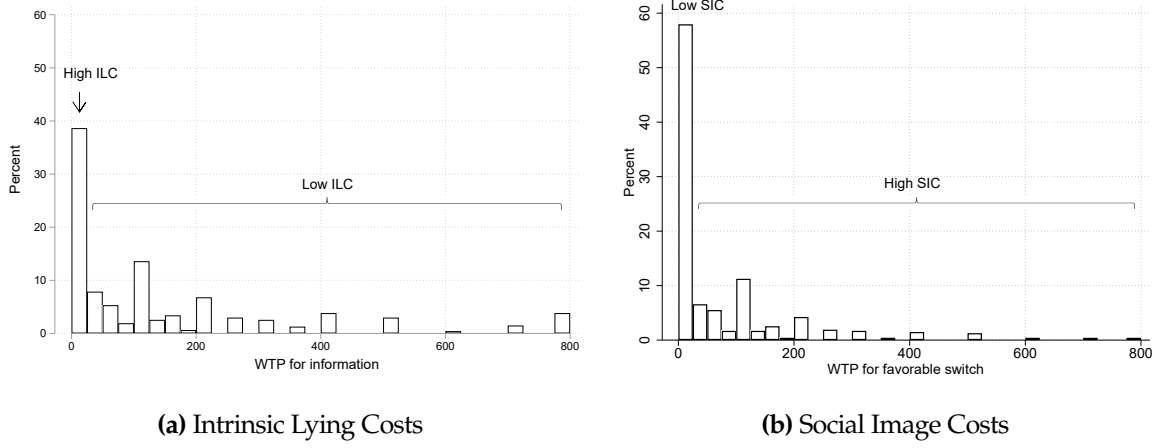


Figure 2: Distributions of ILC and SIC

Notes: Panel a presents the distribution of WTP_{info} , where smaller values imply larger ILC. Panel b presents the distribution of WTP_{switch} , where smaller values imply smaller SIC. General population sample ($n=471$).

costs and social image concerns to matter, they usually refrain from explicitly modeling dependencies of both components. Our novel measure reveals an important empirical insight for the refinement of these models and highlights the potential for additional trade-offs when designing effective policies to reduce dishonest behaviors.

3.2 Preference types

Following our pre-analysis plan, we define types based on whether an individual has a positive WTP_{info} and a positive WTP_{switch} . Note that these types should mainly be interpreted relative to each other, i.e., those who are classified as having low SIC have a lower level of SIC than those whom we classify as having high SIC.²¹ We encounter all four types in our data (see Table 1) and thus identify substantial heterogeneity in IPT.

Table 1: IPT Types

	Low SIC	High SIC	
Low ILC	26.5 %	34.8 %	61.4 %
High ILC	31.4 %	7.2 %	38.6 %
	58.0 %	42.0 %	

Notes: General population sample ($n=471$). A participant is classified as low ILC if their WTP_{info} is strictly larger than zero, and as high ILC otherwise. Conversely, a participant is classified as low SIC if their WTP_{switch} is zero, and as high SIC otherwise.

²¹Our IPT measure also allows for a more fine-grained type classification which could be explored in larger samples. To study the meaningfulness of the current classification, we investigate the behavior of the above-identified types in other decision environments in Section 4.

The most frequent type (34.8%) has low ILC and high SIC. That is, a large fraction of our sample is intrinsically willing to lie but cares at the same time about their social image. Further, we observe a substantial fraction of participants with low SIC and high ILC (31.4%). These participants care about being honest but not about how they are perceived by the Observers. Further, we observe a meaningful fraction of participants who have low ILC and low SIC (26.5%) but only few participants who have high ILC and high SIC (7.2%). Overall, we find systematic heterogeneity in IPT with a substantial fraction of participants who care about only one of the two underlying motivations or none.²²

4 Properties of the measure

4.1 Internal validity

To assess the internal validity of our WTP measures of ILC and SIC and the resulting type classification, we collected additional data on Prolific using two sets of additional treatments that varied the scope to which DMs can incur ILC and SIC in our experimental paradigm. We administered these treatments to show that i) the elicited preference to be honest is sensitive to changes in lying costs, and ii) the elicited preference to be seen as honest is sensitive to changes in social image costs. Further, these treatments allow us to investigate whether we capture these motives independently.

4.1.1 Varying ILC - Design and Results

In our first treatment variation, we randomly assign 499 participants to one of two treatments: one in which ILC play a role (*LieNoObservers*), and one in which ILC play no role (*NoLieNoObservers*). In *NoLieNoObservers*, we set lying costs to zero by instructing participants to choose any item of their liking in the reports. To keep everything else constant, participants generate two items and note them down. After this, they learn that they will be asked to report two items, and that their reports will determine their payoff. They are explicitly instructed that they are allowed to pick any item (from a list of all possible items) in each of their reports. A risk-neutral DM should thus state the expected value of the high-stakes task as their WTP_{info} . As there is nothing to judge in this treatment, we omit Observers and hence the elicitation of social image costs. We compare the *NoLieNoObservers* treatment to *LieNoObservers*, in which we also omit Observers, but keep lying costs as in our standard measure (see Section 2.1).

If the WTP_{info} in our IPT elicitation procedure captures an intrinsic preference for being honest, we will observe that the WTP_{info} is higher in *NoLieNoObservers* as compared to the *LieNoObservers* treatment, and further, that the fraction of participants with a positive WTP_{info} is higher in *NoLieNoOb-*

²²We examine the robustness of type distributions across three different samples in Section 6.1. See also Appendix A.1 for more details on the student sample.

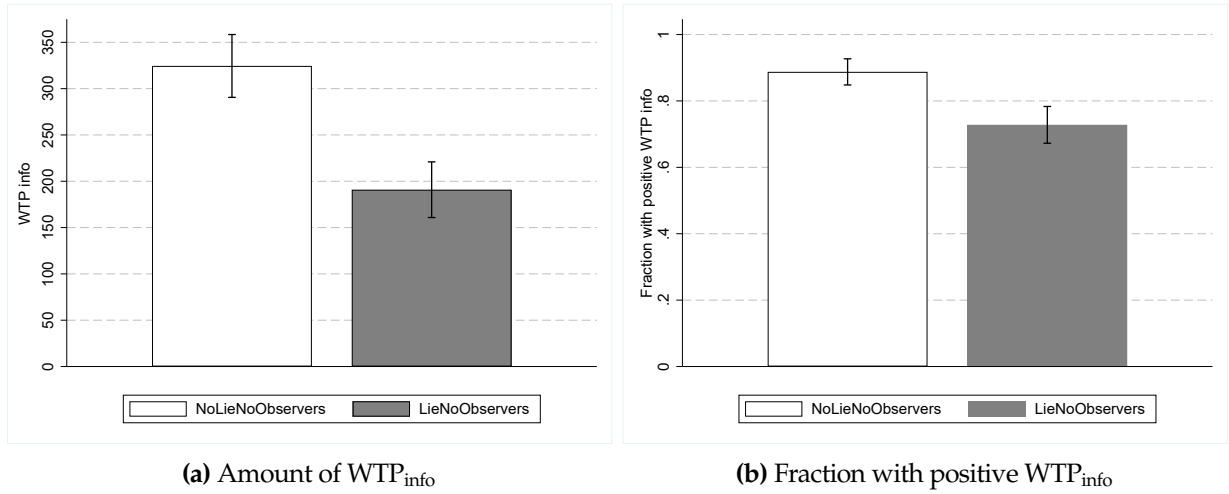


Figure 3: WTP_{info} by ILC treatment

Notes: The figure shows the mean WTP_{info} (in Panel a) and the fraction of participants with positive WTP_{info} (in Panel b) by ILC treatment (*NoLieNoObservers* and *LieNoObservers*). General population sample (n=499).

servers.²³ Figure 3a shows that the mean WTP_{info} in *NoLieNoObservers* in which participants can choose which item to report and in *LieNoObservers* in which participants were asked to report the randomly generated item. The WTP_{info} is indeed significantly higher in *NoLieNoObservers* than in *LieNoObservers* (324.5 vs 190.9, MWU, $p < 0.001$). Figure 3b displays the fraction of participants with a positive WTP_{info} in both treatments. While 72.8% of participants want to know the high-paying item in treatment *LieNoObservers*, this fraction significantly increases to 88.8% in *NoLieNoObservers* (χ^2 -test, $p < 0.001$). Both findings confirm that WTP_{info} meaningfully captures participants' preferences for being honest in *LieNoObservers*.

4.1.2 Varying SIC - Design and Results

In our second treatment variation, we vary whether DMs incur SIC. We randomly assign 503 participants to one of two conditions, in which they are either judged by two Observers (*TwoObservers*) who will learn whether their report resulted in a high but unlikely payoff, or a lower but more likely payoff (in the exact same manner as in our IPT elicitation procedure, see Section 2.1) or not (*ZeroObservers*). Apart from this variation, these treatments are identical to our standard IPT elicitation procedure. Importantly, participants in both conditions learn that they may have either been randomly assigned *TwoObservers* or *ZeroObservers*, which allows us to explain the concept of switching the default task based on which observers may judge the DM's character as well as the WTP_{switch} elicitation procedure (see also Section 2.1) to all participants. Then, before indicating their WTP to switch the default task, DMs learn whether they have been assigned two or zero Observers. If the WTP_{switch} in our standard IPT elicitation procedure captures DMs' SIC, we will observe a higher WTP_{switch} in

²³We pre-registered these hypotheses (AsPredicted No. 131061).

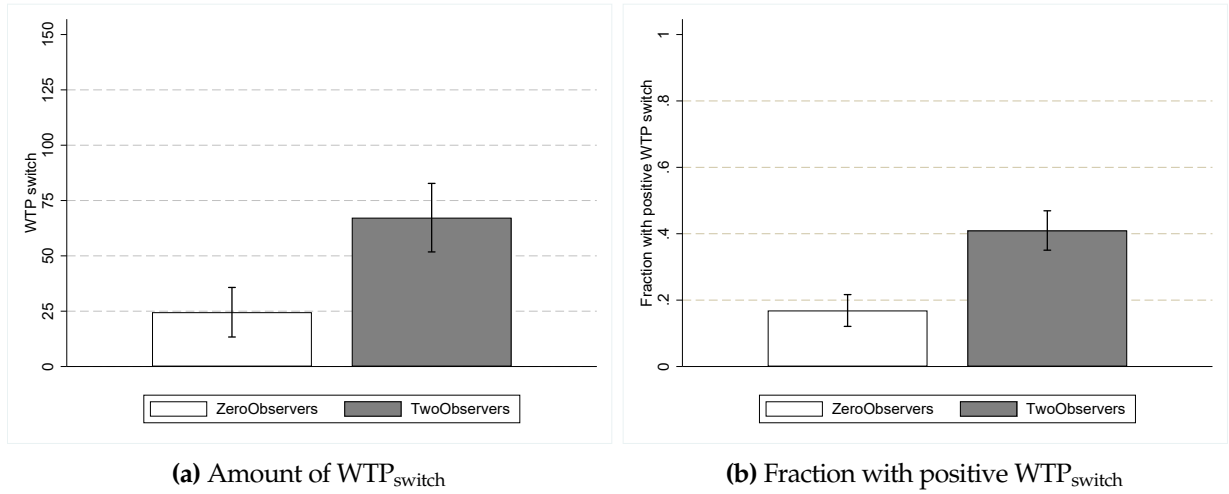


Figure 4: WTP_{switch} by SIC treatment

Notes: The figure shows the mean WTP_{switch} (in Panel a) and the fraction of participants with positive WTP_{switch} (in Panel b) by SIC treatment (*ZeroObservers* and *TwoObservers*). General population sample ($n=503$).

TwoObservers than in *ZeroObservers*; and similarly, the fraction of DMs with a positive WTP_{switch} will be larger in *TwoObservers* as compared to *ZeroObservers*.²⁴

Figure 4a displays the mean WTP_{switch} by treatments. In *ZeroObservers*, WTP_{switch} is significantly lower than in *TwoObservers* (24.5 vs 67.3, MWU, $p < 0.001$). Further, as shown in Figure 4b, the fraction of participants with a positive WTP_{switch} decreases significantly from 41.0% for those randomly assigned to two Observers to 16.9% for those assigned to zero Observers (χ^2 -test, $p < 0.001$). Both findings confirm that being assigned to Observers who judge the character causally increases WTP_{switch} . This WTP hence meaningfully captures preferences for being seen as honest.

Varying whether DMs are judged by Observers or not additionally allows us to test for the independence of preferences for being honest and being seen as honest. To do so, we compare the intrinsic lying costs (i.e., WTP_{info}) of participants who have randomly been assigned zero Observers to those who have been assigned two Observers. Recall that the number of Observers is known before participants are asked to state their WTP_{info} : If the WTP_{info} in *TwoObservers* does not differ from the WTP_{info} in *ZeroObservers*, the WTP_{info} captures ILC independently of SIC.²⁵ Indeed, we find that the WTP_{info} does not differ significantly between those with two and zero observers (165.0 vs 165.9, MWU, $p=0.612$) and further, that also the fraction of participants with a positive WTP_{info} does not differ significantly (67.1% in *ZeroObservers* vs. 69.2% in *TwoObservers*, χ^2 -test, $p=0.616$). Hence, we conclude that our IPT paradigm independently measures preferences for being honest and preferences for being seen as honest.

²⁴We pre-registered these hypotheses (AsPredicted No. 129232).

²⁵This hypothesis builds on the idea that the elicitation of the WTP_{info} for the favorable switch allows participants to eliminate potential SIC with respect to the observers and was pre-registered at AsPredicted (No. 129232).

4.2 Predictiveness

To assess whether the IPT measure meaningfully relates to (dis)honest behavior in other environments, we next discuss its relationship to behavior in three other experimental paradigms that allow for dishonest acts. The first paradigm is a mind game, in which the true state is impossible to observe for the experimenter and in which payoffs are similar to the widely-studied die-rolling paradigm by Fischbacher and Föllmi-Heusi (2013). The mind game reflects a class of reporting paradigms that are close in nature to our IPT measure: the true state of the world is unobserved and the report entails no explicit negative externalities for other participants. As such, the mind game – as other unobserved reporting tasks – abstracts from social preferences and strategic interactions thereby capturing potential ILC and SIC jointly (Abeler et al. 2019).

The second paradigm relates to a widely used set of paradigms in which the dishonesty of an informed sender may harm an uninformed receiver of a message. These sender-receiver games were introduced first by Gneezy (2005). In contrast to unobserved reporting paradigms, misreporting for the own benefit can be observed by the experimenter and imposes a negative externality on the matched partner. In addition to ILC and SIC, behavior may hence be affected by social preferences. We design a variant of the sender-receiver game that is non-strategic from the perspective of the sender (Gneezy et al. 2013). We administer our sender-receiver game in two conditions that vary the social image costs vis-à-vis the receiver but keep social preferences for this receiver constant. This allows us to study to what extent SIC measured in the IPT task based on a third-party observer is predictive of SIC regarding an affected party, holding ILC constant.

In both paradigms, the mind game and the sender-receiver game, participants can misreport a (random) state of the world to increase their material benefit. This rules out any social image costs other than being seen as honest. However, in many situations, individuals do not lie to increase their payoff, but rather to improve their social image related to, e.g., their performance or knowledge (Barron et al. 2022; Hugh-Jones 2016). In these situations, social image concerns may increase misreporting rather than decrease it. We study whether SIC measured in the IPT elicitation procedure relate to behavior in a novel paradigm, in which participants may lie to improve their social image by appearing knowledgeable in a domain they care about absent material benefits.

As running these additional paradigms in combination with the IPT measure within the same session is more time-intensive, we administered these tasks in the student sample. Participants encounter the three additional paradigms in randomized order and we elicit the IPT measure either before or after the other paradigms, which we describe in more detail below.²⁶ Note that, we included the mind game also in the Prolific sample (as it is relatively short), and obtain very similar results as in the student sample (see Appendix Section B.1).

²⁶Pre-registered on AsPredicted No. 70550.

4.2.1 Mind game

Design We design a variant of a mind game (Greene and Paxton 2009; Jiang 2013; Kajackaite and Gneezy 2017). In this task, we ask participants to pick and memorize one out of six African cities from a given list. On the next screen, participants learn the association between the reported city and payoff (ranging from 0 to 1250 ECU, in steps of 250 ECU), and make their report.²⁷ As such, the payoffs are similar in structure to the widely-used die roll paradigm (Fischbacher and Föllmi-Heusi 2013) and the potential size of a lie corresponds to the distance in monetary payoffs of the reported as compared to the observed city. We expect that intrinsic lying costs influence the claimed monetary payoff, i.e., those with lower intrinsic lying costs claim larger payments. In the IPT measure, low intrinsic lying costs are reflected in high WTP_{info} . We thus hypothesize that WTP_{info} correlates positively with payoff in the mind game. Further, we suspect that SIC (reflected in high WTP_{switch}) may reduce the likelihood of reporting the highest payoff in the mind game conditional on having low ILC.

Results Correlating WTP_{info} and the payoff in the mind game, we find that the two are positively and statistically significantly related (Spearman's $\rho = 0.160$, $p=0.004$). This is in line with our hypothesis and implies that those with lower intrinsic lying costs claim larger payoffs.²⁸ Regarding social image concerns, we find no statistically significant relationship between WTP_{switch} and the payoff in the mind game (Spearman's $\rho = -0.065$, $p=0.240$). The correlation is of similar magnitude when focusing on the relationship between reporting the maximum amount possible in the mind game and the social image costs (Spearman's $\rho = -0.069$, $p=0.211$). However, for participants who intend to lie in the IPT paradigm ($WTP_{info} > 0$), we find a statistically significant negative relationship between reporting the maximum amount possible in the mind game and their SIC identified in the IPT measure (Spearman's $\rho = -0.122$, $p=0.068$) whereas for those who have high intrinsic lying costs according to the IPT measure ($WTP_{info} = 0$) image concerns seem not to matter for their decision to report the maximum amount possible in the mind game (Spearman's $\rho = -0.001$, $p=0.997$).

Turning to the type classification, we first compare the average payoffs in the mind game of those whom we classify as low ILC types (987.7 ECU) vs those whom we classify as high ILC types (876.2

²⁷ Associations between cities and payoffs are randomized at the individual level. We chose African cities, because we considered it less likely that our European participants would have personal preferences or connections to any of the cities compared to if we had used European cities for example. Using these cities rather than numbers renders it also less likely that our participants incur additional lying costs due to some natural distance between the items to be reported (Gneezy et al. 2018).

²⁸ This correlation is robust in magnitude and significance to a parametric assessment that controls for risk aversion and curiosity (see Columns 1 and 2 in Table A.3). The regression coefficients of the risk and curiosity measures are statistically indistinguishable from zero, when we regress WTP_{info} and/or WTP_{switch} on the extent of potential cheating behavior in the mind game. Further, the coefficient of WTP_{info} remains robust in terms of size and significance across various specifications. Similarly, controlling for socio-demographics (gender, income, age categories and political orientation) does not reduce the predictive power of IPT. See Appendix Section A.3 for a detailed discussion of the robustness of results. Further, we also observe a positive correlation between WTP_{info} and the payoff in the mind game in the general population sample (see Appendix Section B.1).

ECU). This difference is statistically significant (MWU, $p=0.010$) and implies that low ILC types report significantly higher payoffs than high ILC types. However, as Fischbacher and Föllmi-Heusi (2013) note, reporting in such tasks can also be affected by social image costs. In particular, individuals may not misreport to the full extent to appear more honest, hence ‘disguising their lie’. We therefore show the full distribution of claimed payoffs separately for our four IPT types in Figure 5. The left panels present reports by types with high ILC, while the right ones show reports by types with low ILC. The top row displays reports by types with low SIC, while the bottom row shows reports by types with high SIC. The solid horizontal line represents 16.7%, i.e. the expected fraction of reports if all participants are fully honest.

We first examine the behavior of participants with low SIC (top row of Figure 5). Clearly, the distribution of reports differs for DMs with high ILC (top, left) as compared to DMs with low ILC (top, right) when social image costs are low. With low ILC, a larger fraction of DMs reports the highest (χ^2 -test, $p=0.026$) and DMs participants report lower payoffs (i.e., the distribution shifts to the right, MWU, $p=0.001$). At first glance, it may be surprising that not all participants with low ILC report the highest payoff as we examine participants with low SIC. Note however, that those classified as low ILC differ in their WTP_{info} (see Figure A.1a for the distribution in the student sample), such that some participants may find it too costly in terms of their ILC to lie to the full extent.

Since types are not equally distributed (see also Table A.1), the number of observations is different in each panel and is particularly low for those with high ILC and high SIC (bottom, left). This makes it difficult to compare this panel to the others. To examine the effect of social image costs, we hence focus on the comparison of those with low ILC (top, right vs bottom, right). While participants in both panels have low ILC, the additional effect of SIC becomes visible in a smaller share of highest payoffs (χ^2 -test, $p=0.029$). Further, we observe a shift of the distribution of payoffs to the left (MWU, $p=0.008$).

4.2.2 Sender-receiver game

Design In the sender-receiver game, a receiver chooses between different options without knowing the payoffs associated with the options. To avoid sophisticated forms of deception by telling the truth (see Sutter 2009), we let receivers choose one out of ten available options. One option results in a high payoff of 900 ECU while the other nine options result in a low payoff of 100 ECU for the receiver. The sender advises the receiver by sending a message of the form “Option X yields the highest possible payoff for you.”, which can be truthful or deceptive. In our version of the paradigm, the sender’s payoff only depends on the message they send, with the truthful message yielding a low payoff of 100 ECU, and a deceptive message yielding 900 ECU. Using the strategy method, the sender chooses the message in two conditions, varying whether lies are deniable or not. In Condition 1, the receiver does not know which possible payoff levels exist. The only information they receive is the message

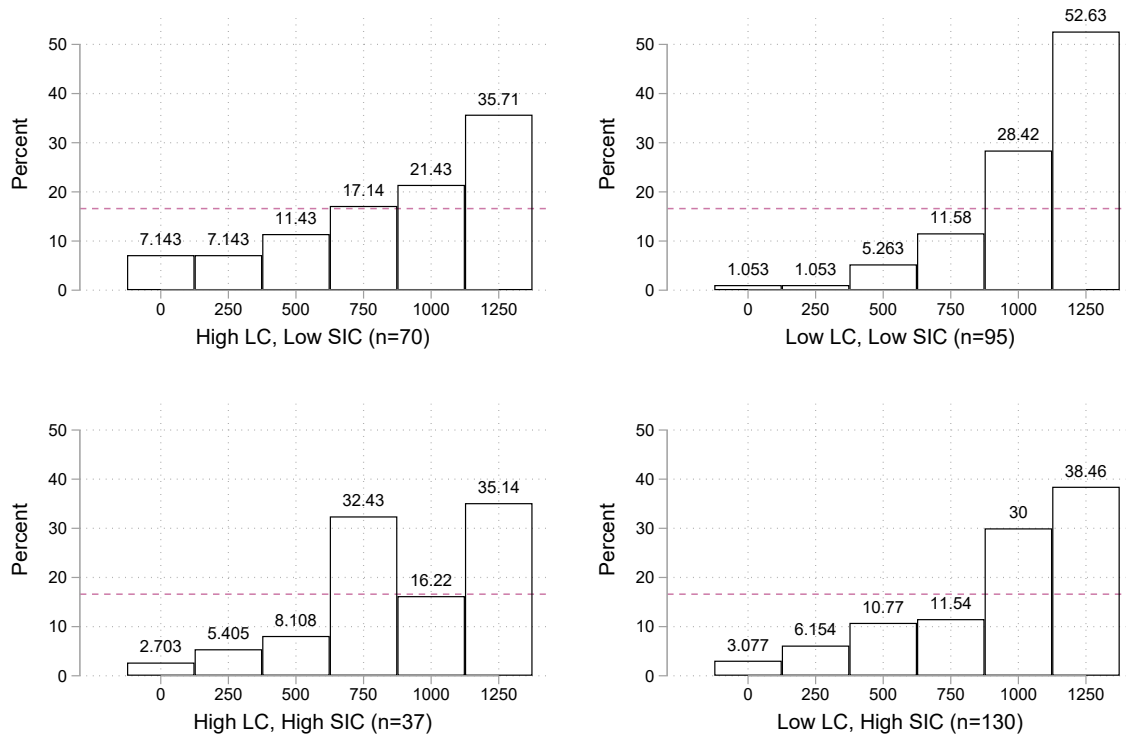


Figure 5: Claimed payoffs in the mind game by IPT types

Notes: The figure shows the distribution of claimed payoffs in the mind game by IPT types in the student sample ($n=331$). *ILC* denotes intrinsic lying costs and *SIC* denotes social image costs. The solid horizontal line represents the expected fraction of reports for each payoff when all participants report honestly.

from the sender. Thus, receivers do not learn whether a message was true or not in this condition (i.e., a lie is deniable). In Condition 2, the receiver is informed that one option yields a payoff of 900 ECUs while the nine other options yield a payoff of 100 ECUs and, as in Condition 1, receive a message from the sender. Thus, receivers can infer whether the message was true or not (i.e., a lie is non-deniable).

The receiver is unaware of the two possible conditions and only receives information based on the randomly determined payoff-relevant condition. Further, the receiver does not learn the sender's payoff in either of the conditions. The two main outcome measures of this task are the type of message sent (truthful or deceptive) in the two conditions (deniable vs. non-deniable). As we are interested in DMs' behavior as a sender, we inform senders that the computer randomly determines whether their message will be delivered (or not) before the other participant chooses an option. This allows us to allocate a large fraction of DMs (319 subjects) to the role of the sender, and deliver a randomly selected subset of messages to the remaining DMs (12 subjects) in the role of receivers.²⁹

We expect that those with lower intrinsic lying costs are more likely to send a deceptive statement, and thus hypothesize that WTP_{info} correlates positively with the payoff in the sender-receiver game. Further, we expect that senders who have higher social image costs (according to the IPT measure)

²⁹See also the instructions in Appendix Section E.2.2.

are more likely to react to deniability (i.e., to the difference across conditions in the sender-receiver game). Hence, we expect high SIC senders to be more likely to send a deceptive message when lies can be denied, but to refrain from doing so when lies can be identified by the receiver.

Results Each sender makes two decisions: one message choice with higher social image implications in the non-deniable condition and one message choice with lower social image implications in the deniable condition. Externalities are constant across these two decisions. 17.2% of the senders react to the two choices in the expected way and only send a deceptive message when social image implications are low. Based on their behavior in the sender-receiver game, we classify these individuals as caring about their social image.³⁰

Following our hypotheses, we first analyze whether the payoff in the sender-receiver game positively correlates with WTP_{info} . Contrary to this hypothesis, we do not find a positive correlation, neither in the deniable condition (Spearman's $\rho = 0.086$, $p = 0.126$), nor in the non-deniable condition (Spearman's $\rho = 0.044$, $p = 0.430$). Since lying is observable at the individual level, we additionally study the predictiveness of the WTP_{info} for participants who have no SIC in the IPT measure (i.e. those who may not mind being perceived as dishonest). Indeed, we find a positive correlation of the WTP_{info} and the payoff in the sender-receiver game for this sub-sample in the deniable condition (Spearman's $\rho = 0.170$, $p = 0.034$) and in the non-deniable condition (Spearman's $\rho = 0.158$, $p = 0.048$). Next, we test whether there is a relationship between the measures of social image costs in IPT and reactions to deniability in the sender-receiver game. For participants who react to deniability in the sender-receiver game, the average WTP_{switch} is equal to 248.18, while for participants who do not react to deniability in the sender-receiver game, the WTP_{switch} amounts to 171.21 (MWU, $p = 0.018$). In addition, we separately examine behavior in the sender-receiver game for the low SIC types and the high SIC types. Figure 6 shows that participants with high SIC in the IPT task are more likely to react to the variation and only lie in the deniable condition as compared to participants with low SIC in the IPT task (χ^2 -test, $p = 0.041$). In line with our hypothesis, we find that being classified as having SIC in the sender-receiver game correlates positively with WTP_{switch} (Spearman's $\rho = 0.114$, $p = 0.042$). We additionally find that those with high SIC in IPT are less likely to lie in both conditions (χ^2 -test, $p < 0.001$). We conclude that our measure of SIC in the IPT task meaningfully predicts behavior related to social image concerns in sender-receiver games.

³⁰We chose the parametrization to roughly match expected payments from IPT, as only one of the four tasks (IPT and the three validation tasks) was randomly determined to be payoff relevant. We note that this likely induced a majority of participants to send two deceptive messages (62.4%), such that we unfortunately cannot assess their social image costs in this task. Only 16.3% percent choose the honest message in both conditions.

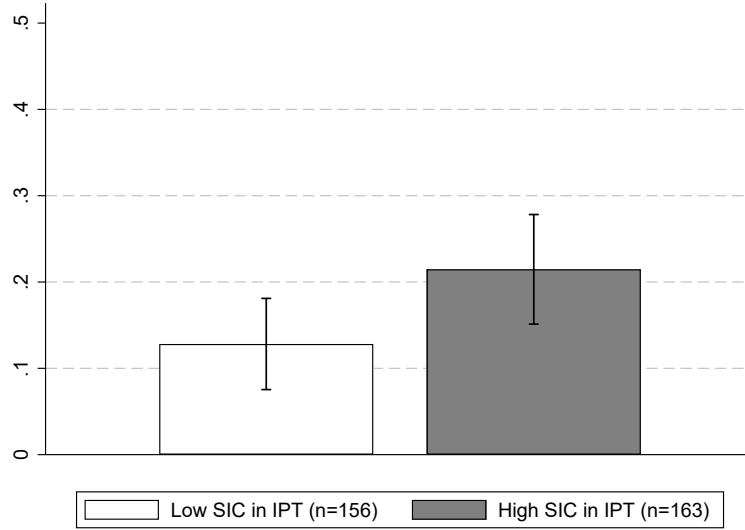


Figure 6: Fraction with SIC in sender-receiver game by SIC in IPT

Notes: The figure shows the fraction of senders that lie in the deniable condition but not in the non-deniable condition, that is, those whom we classify as having social image costs (SIC) in the sender-receiver game, separately for low SIC and high SIC types measured by IPT. Student sample (n=319).

4.2.3 Knowledge reporting task

Design Finally, we measure participants' preferences for truth-telling when the benefits from dishonest behavior are non-monetary and can solely enhance participants' social- or self-image. To do so, we design a knowledge reporting task in which participants are shown two lists, one after the other, containing 20 entries each. They are asked how many entries in the list they identify as either important people in human history (list A) or birds (list B).³¹ Both lists contain six entries almost everyone knows and 14 entries that do not correspond to important people or existing birds.³² Decision-makers are informed that the number of identified entries is shown, together with the decision-makers' identifier, to independent Observers (those from IPT measure) at the very end of the experiment.³³ Participants receive a flat payment of 500 ECU for entering the identified number of humans (birds, respectively) for each list, i.e., there is no monetary incentive to misreport.³⁴

In a pilot with 32 participants from the same student subject pool, we incentivized participants to choose the existing important people in human history and real birds from the lists. On average, participants who provided an answer (and did not time out without answering) identified 6.3 humans and 6.3 birds. In addition, participants identified the same number of humans and birds (Signrank test, $p=0.700$). We are thus confident that, on average, participants knew that there were 6 real humans and

³¹The order in which lists are displayed is randomized at the individual level.

³²See E.2.3 for the list of people and birds.

³³This task was inspired by the work of Trocinska (2020).

³⁴We impose a 60-second time limit per list to avoid searching for answers online. If participants enter the number of entries they identify as important people or birds within the time limit, they receive the payment. We exclude 19 participants who time out for one or both lists from our analysis. This leaves us with 312 observations.

6 real birds in the lists. The pilot data is also useful to provide an internal validity check for our task. We designed this task to study whether people are willing to lie for their social image, i.e., to appear knowledgeable, independent of financial incentives. In the main experiment, participants report having identified 7.4 humans and 7.8 birds, on average. These numbers are significantly larger than in the pilot (MWU tests, $p < 0.001$ for humans and birds), confirming that participants are on average willing to lie for a positive social image if this bears no costs. To capture the importance of appearing knowledgeable in the two tasks, we further asked participants to rate how comfortable another participant would feel if they had identified 5 important people in human history (birds respectively), on a 7-point Likert scale from 1 (very uncomfortable) to 7 (very comfortable).³⁵ We observe that the number of reported identified humans (birds) correlates with the comfortableness ratings (humans: Spearman's $\rho = -0.319$, $p < 0.001$, birds: $\rho = -0.260$, $p < 0.001$), i.e., the less comfortable participants think someone else would feel when reporting only 5 items, the more items they report. Based on these analyses, we are confident that this new task captures social image concerns in the absence of monetary incentives.

To analyze whether our IPT measure predicts behavior in this task, we construct a variable for social image concerns in the knowledge reporting task. Based on the comfortableness ratings, we first assess on the individual level whether humans or birds are considered more important to know. We then subtract the reported number of identified items in the less important task from the reported number of identified items in the more important task. Thus, if a participant considers birds more important than humans, we subtract the number of identified humans from the number of identified birds and vice versa, such that a positive number indicates social image concerns. We hypothesize that this variable correlates positively with WTP_{switch} .³⁶ We further hypothesize that for those with positive WTP_{switch} , the number of reported entries in the task considered more important to be knowledgeable in will correlate positively with WTP_{info} . The reason is that lying is only relevant for those with SIC in this task given that there is no monetary incentive to lie. For those who have an incentive to misreport because they have positive SIC, we expect that the extent of lying depends on ILC.³⁷

³⁵We presume that there is a high correlation of between DMs' expectations about how comfortable another participant feels and how comfortable they feel themselves and thus opted for this approach to avoid that DMs will adjust their rating based on their own reports.

³⁶We originally assumed that social image costs are larger for not knowing important people in history rather than for not knowing birds. We thus preregistered our outcome variable of interest as being the difference in entries between humans and birds (humans - birds) and hypothesized that this would positively correlate with WTP_{switch} . We find no support for this hypothesis; the correlation is close to zero and not statistically significant (Spearman's $\rho = 0.010$, $p = 0.860$). The original assumption of appearing more knowledgeable regarding humans being more important for the social image is, however, only true for 33.3% of our participants. 13.5% consider knowing birds more important, and 53.2% do not distinguish between the two. Given that our original assumption is only true for a small fraction of participants, we deviate from our pre-analysis plan by taking into account which items participants consider more relevant for the social image. Following the spirit of our pre-analysis plan, we continue to focus on the difference in reported items. Correlating the difference in reported items with the difference in comfortableness ratings reveals a significant relation (Spearman's $\rho = -0.248$, $p < 0.001$), again supporting the internal validity of our task.

³⁷Originally, we hypothesized that the number of entries in the humans' task would correlate positively with WTP_{info} for those with positive SIC, for the same reason stated in the previous footnote. We again deviate due to the underlying assumption being invalid. The original hypothesis finds no support (Spearman's $\rho = 0.068$, $p = 0.393$)

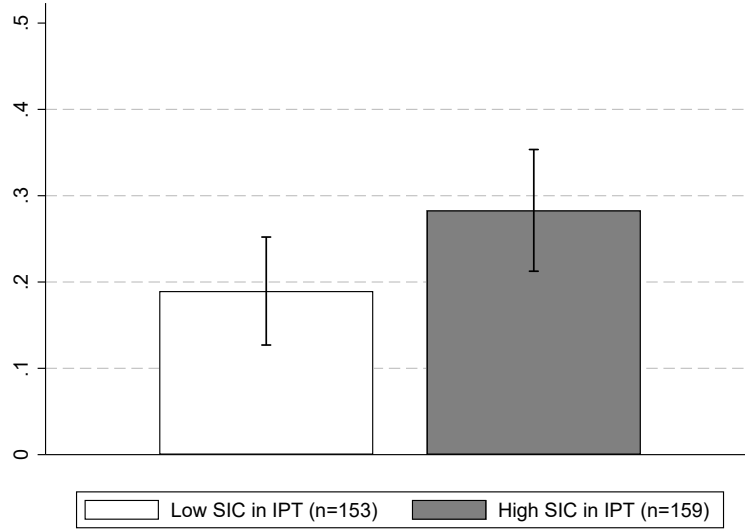


Figure 7: Fraction with SIC in knowledge reporting task by SIC in IPT

Notes: The figure shows the fraction of participants classified as having SIC in the knowledge reporting task (reported more entries in the task they consider more important) and 95%-confidence intervals, separately for low social image costs (SIC) and high SIC. Student sample (n=312).

Results We analyze whether our IPT measure predicts behavior in the knowledge reporting task, and correlate the difference in reported items between the task considered more and less important, respectively, with WTP_{switch} . We find a positive correlation that is marginally significant (Spearman's $\rho = 0.140$, $p = 0.091$). We further examine the extensive margin and analyze whether the fraction of participants classified as having SIC in the knowledge reporting task, that is participants who report more items in the task they consider more important, differs according to our classification in IPT. Figure 7 shows the fraction of participants classified as having social image costs in the knowledge reporting task by having low or high SIC in the IPT task. The probability of being classified as having social image concerns in the knowledge reporting task is marginally significantly different between high SIC (28.8%) and low SIC (19.0%) types as classified in IPT (χ^2 -test, $p = 0.052$). Further, we find a significantly positive correlation between SIC in the knowledge reporting task and SIC in IPT (WTP_{switch} ; Spearman's $\rho = 0.115$, $p = 0.044$). Finally, we analyze whether the difference in reported items is correlated with WTP_{info} for those with SIC in IPT. This correlation is weakly positive (as expected) but statistically insignificant (Spearman's $\rho = 0.119$, $p = 0.285$).

4.3 Further properties

As it is also important to understand how a new measure can be used in practice, we also studied whether it matters when we elicit IPT. We randomly conducted the IPT elicitation procedure in the student sample either at the beginning or at the end of the experimental session. That is, the IPT measure was elicited either before or after three other paradigms that capture certain aspects of

honesty preferences.³⁸ Reassuringly, we find no order effects: Elicited WTPs are robust to administering the measure at the beginning of a session or after the three validation tasks (MWU tests, $p=0.301$ for WTP_{info} , $p=0.937$ for WTP_{switch}). Further, the distribution of classified types does not differ depending on whether IPT is administered at the beginning or at the end of a session (χ^2 -test, $p=0.998$). As the additional tasks also concern moral behavior, this implies that the IPT measure is not only robust to order effects but also to potential moral licensing across tasks.

5 Survey measure

5.1 Design

In addition to the incentivized IPT measure, we developed a short survey module that proxies preferences for being and being seen as honest at the individual level. This measure can be used when a quick, non-incentivized measurement of preferences is desirable. Akin to our incentivized measure, we aim for one context with two independent decisions. In contrast to the experimental measure, in which we measure a range of ILC and SIC with WTPs, the short survey module focuses only on a binary classification, i.e., higher and lower intrinsic lying costs and higher and lower social image costs. The module consists of a vignette and two questions. The vignette describes a situation in which the participant is called by the host of a live radio show to participate in a raffle. The task in the raffle is to flip a coin four times and report the number of flipped ‘tails’, with each reported tail yielding a payoff of 10 currency units (e.g. euros, pounds, dollars). Participants are assured that the host has no way of verifying the reports, and are then asked how many ‘tails’ they would report had they in fact flipped 0 tails (Question 1) or 4 tails (Question 2).³⁹

Those two questions are designed such that they proxy preferences for being honest (Question 1), and being seen as honest (Question 2). In Question 1, reporting 0 tails (when in fact 0 tails were observed) is likely motivated by high intrinsic lying costs that outweigh the monetary gains from lying.⁴⁰ Hence, we classify participants who report 0 tails in Question 1 as having relatively high intrinsic lying costs, and those who report a positive number as having relatively low intrinsic lying costs. In Question 2, not reporting 4 tails cannot be driven by preferences for being honest but only by preferences for being seen as honest. We hence classify a report of 4 tails in Question 2 as indicative of relatively low social image costs and a report of a smaller number of tails as indicative of relatively high social image costs.

³⁸The three additional tasks were used to assess the predictiveness of our IPT measure (as explained above).

³⁹The exact phrasing can be found in Appendix E.

⁴⁰In principle, participants with preferences for being seen as honest and low intrinsic lying costs may also be motivated to report a positive number of ‘tails’, namely the most likely outcome of 2 times ‘tails’, but the latter is empirically rarely observed.

5.2 Correlation with the experimental IPT measure

To experimentally validate our survey measures, we administered the survey module in the convenience sample on Prolific ($n=471$, for details see Section 2.2). We analyze whether our type classification based on the survey questions corresponds to the classification based on the experimental IPT measure. We start with intrinsic lying costs. Figure 8a displays the fraction of participants reporting zero tails when the true outcome is zero, separately for those participants with high intrinsic lying costs (based on the experimental measure, ILC_{ex}) and those with low ILC_{ex} . Among those with high ILC_{ex} ($n=182$), 76.4% do not lie and report zero tails if the true number is zero. This fraction decreases to 50.5% for those with low ILC_{ex} ($n=289$), which is significantly smaller than in the group with high ILC_{ex} (χ^2 -test, $p<0.001$). Similarly, the WTP_{info} is significantly smaller for those who report zero tails compared to those who report more (90.6 vs 205.1, MWU test, $p<0.001$).⁴¹

Next, we analyze the survey question aimed at eliciting SIC. Figure 8b shows the fraction of participants reporting a number of tails smaller than 4 when the true number is 4, by their experimental social image cost classification. Among those with low SIC_{ex} ($n=273$), 14.7% report less than 4 tails if 4 is the true number, while among those with high SIC_{ex} ($n=198$), 25.8% do. Hence, those with high SIC_{ex} are more likely to also exhibit SIC in the survey measure (χ^2 -test, $p=0.003$). Similarly, the experimental social image costs (WTP_{switch}) are significantly larger for those reporting less than 4 ($n=91$) than for those who report 4 ($n=380$) in the survey (84.0 vs 59.5, MWU test, $p=0.002$).⁴²

Finally, we examine to what extent the two type classifications based on the experiment and the survey measure overlap. For 41.8% of our sample, the two distinct measures yield identical classifications. This alignment significantly exceeds what would be expected if types were random. Nonetheless, the overlap is far from complete, a finding that is unsurprising given the inherent differences in the tasks involved. First, the experimental measure incorporates incentives, while the survey measure does not. This may explain why the fraction of individuals classified as having high intrinsic lying costs is different for the survey measure (60.5% vs 38.6% in the experimental measure). This larger fraction with high ILC is in line with more honesty in unincentivized reports (Charness et al. 2019) or hypothetical questions (Shalvi et al. 2011). Second, in the experimental measure, the true state remains unobservable to the experimenter, whereas it is explicitly provided in the vignette. More critically, the role of the Observers in the incentivized measure is more pronounced compared to that of the radio show host, as the Observers explicitly evaluate the decision-makers' (DMs') choices, while no such evaluation is implied in the vignette. This distinction may explain why a smaller

⁴¹The WTP is also smaller for those who report zero tails when restricting the sample to participants with a positive WTP_{info} (176.9 vs 266.8, MWU test, $p<0.001$).

⁴²The observed differences are driven by the extensive margin (i.e., by whether participants are willing to buy the favorable switch in the IPT task). When restricting the sample to participants with a positive WTP_{switch} , we do not find a statistically significant difference in WTP s for those who report four tails (155.1) and those who do not report four tails in the survey question (150.5, MWU test, $p=0.856$).

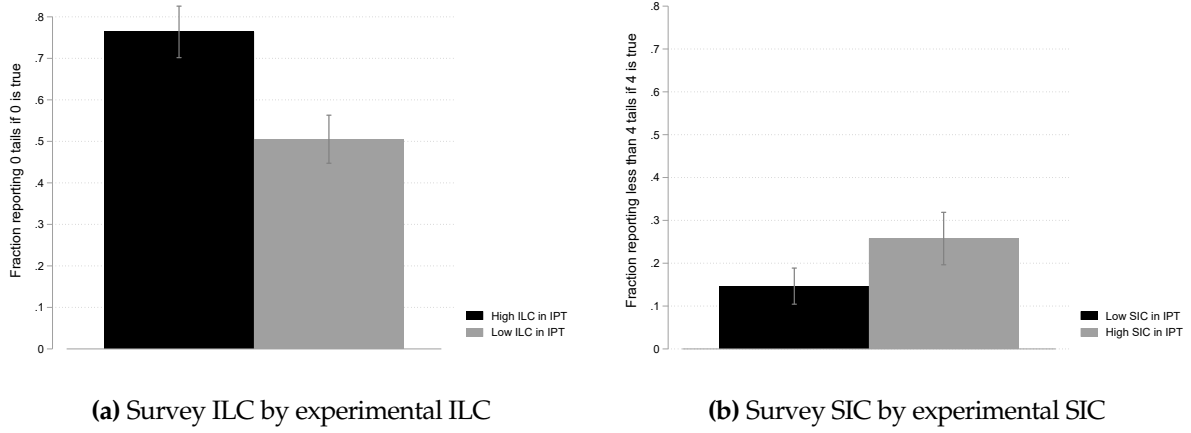


Figure 8: Survey types by experimental IPT types

Notes: General population sample ($n=471$). Panel (a) presents the fraction of participants reporting honestly zero tails in Question 1 (survey measure for intrinsic lying costs, ILC), and 95% confidence intervals, separately for high ILC_{ex} and low ILC_{ex} types. Panel (b) presents the fraction of participants underreporting the number of tails in Question 2 (the survey measure for social image costs, SIC), and 95% confidence intervals, separately for high SIC_{ex} and low SIC_{ex} types.

proportion of individuals are classified as having high social image costs in the survey (19.3% vs. 42.0% in the experimental measure). Irrespective of the differences in the identified type distributions across elicitation methods, the systematic differences in WTP_{info} and WTP_{switch} by the types classified with the survey measure indicate that the survey measure can serve as a useful proxy for capturing both motives when an incentivized procedure is impractical. The latter is also confirmed by the predictiveness of the survey module presented below.

5.3 Predictiveness of survey types

We assess the predictiveness of the type classification based on the survey measure using the mind game.⁴³ We analyze the data analogously to Section 4.2.1. Those with low ILC (i.e., they report more than zero tails when the true outcome is zero) claim larger payoffs (853 vs. 714 points; MWU, $p<0.001$). This is also evident in Figure 9. Comparing its upper two panels, we note that the percentage of participants reporting the maximum payoff doubles when moving from those classified as having high ILC to those having low ILC (holding SIC constant; χ^2 , $p<0.001$).

Regarding social image costs, and similar to the predictiveness with the incentivized IPT measure (presented in Section 4.2.1), we find no statistically significant difference in claimed payoffs (MWU, $p=0.342$). Focusing on the fraction claiming the highest possible payoff, we also find no difference between those with high and low SIC (χ^2 , $p=0.104$). However, for participants with low intrinsic lying costs (right panels in Figure 9), we find that the fraction claiming the highest possible payoff is significantly lower for participants with high SIC (18% vs 44%, χ^2 , $p<0.001$). For participants with

⁴³Further evidence on the predictiveness of our survey measure in the context of a sender-receiver game is provided by Feess et al. 2024.

high ILC (left panels in Figure 9), social image costs do not appear to matter in the decision to claim the highest possible payoff in the mind game (20% vs 23%, χ^2 , $p=0.754$). These findings align well with the predictiveness of the incentivized IPT measure, suggesting that the two survey questions can meaningfully serve as proxies for ILC and SIC.

5.4 Implementation in a panel study

In this section, we briefly showcase the potential value of the survey module by exploring the relationship between individual preferences for truth-telling and a variety of economic decisions and outcomes.⁴⁴ To be able to do so, we introduced the two survey items capturing ILC and SIC (in random order) in the Innovation Sample of the German Socioeconomic Panel (2023 SOEP-IS).⁴⁵ The SOEP-IS is designed for innovative data collection and thus ideal for developing and testing new measurement tools. The 2023 SOEP-IS data set is comprised of an approximately representative sample of the

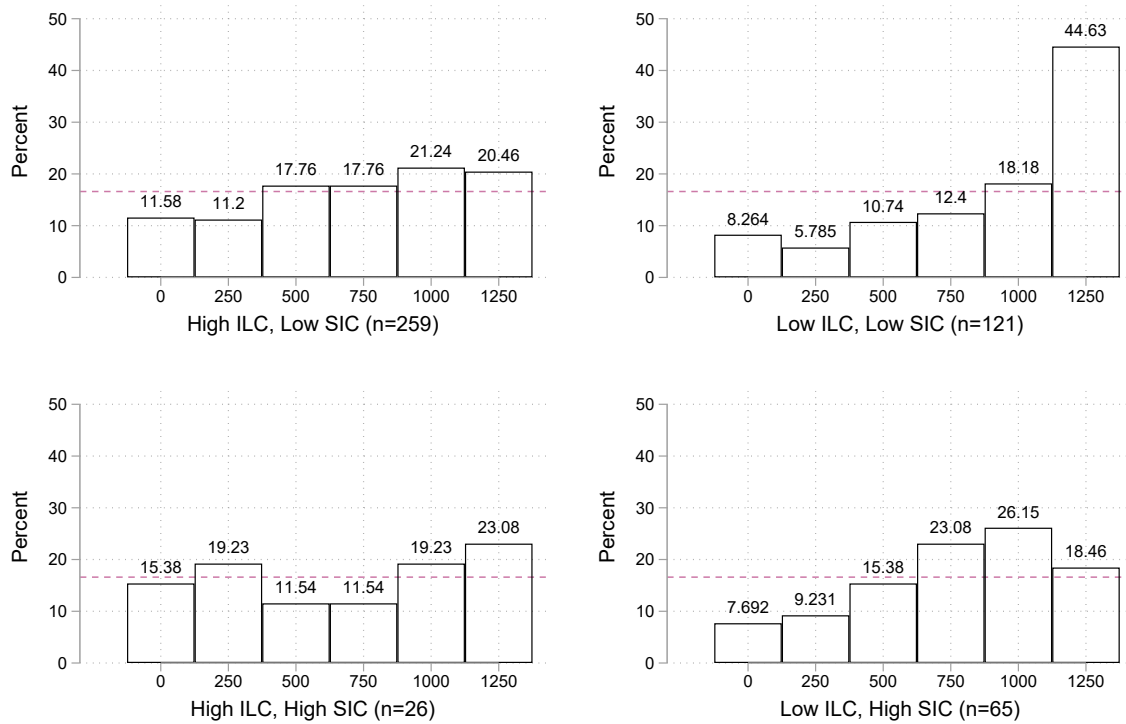


Figure 9: Claimed payoffs in the mind game by types (based on survey measure)

Notes: The figure shows the distribution of claimed payoffs in the mind game by types (defined based on the survey measure) in the general population sample ($n=471$). *ILC* denotes intrinsic lying costs and *SIC* denotes social image costs. The dashed horizontal line represents the expected fraction of reports for each payoff when all participants report honestly.

⁴⁴Note that the survey measure can also be successfully used as an effective control variable, see for example Feess et al. (2024).

⁴⁵Participants answered the questions themselves to avoid potential image concerns related to the interviewer.

German population (2,387 persons).⁴⁶ The SOEP-IS is part of the German Socioeconomic Panel - one of the largest and longest-running multidisciplinary household surveys worldwide – such that we are able to link (a subset of) SOEP-IS participants to their responses in previous SOEP waves.

While numerous economic decisions and real-life outcomes may be linked to individual preferences for truth-telling, we preregistered to explore the role of ILC and SIC across three distinct dimensions: labor market decisions and outcomes, household formation, and the influence of early life experiences on preferences for being, or being seen as, honest. Based on previous literature, we developed hypotheses on the relationship between real-world outcomes and ILC / SIC and preregistered them before gaining access to the 2023 SOEP-IS data. It is important to note that at the time of preregistration (see: <https://osf.io/fqswb>), we were unaware of the specific participants who would be involved in the 2023 SOEP-IS data collection. Consequently, we could not be certain whether the newly collected sample would sufficiently overlap with previous waves to enable the testing of our preregistered hypotheses. Fortunately, this overlap occurred for almost all variables of interest.⁴⁷

5.4.1 Labor Market: Selection, Outcomes, and Dynamics

Previous literature suggests that preferences for truth-telling affect labor market decisions and labor market outcomes. Depending on whether individuals have a preference for being honest, they may self-select into decision environments in which their preference type benefits the most (Fehrler et al. 2020b; Saccardo and Serra-Garcia 2023). For instance, it has been shown that corruption propensities among Indian public sector aspirants are higher than among private sector aspirants (R. Banerjee et al. 2015) and that Indian university students' willingness to enter public service is higher for dishonest students (Hanna and Wang 2017). In contrast, in less corruption-prone public environments, positive selection into public service has been observed. For example, Barfort et al. (2019) found that the more honest Danish individuals were, the more likely they wished to work in public service.⁴⁸ Based on these findings, we hypothesized that individuals classified as having low intrinsic lying costs have a different likelihood to work as a public servant in Germany (as compared to individuals classified as having high ILC).

The first two columns of Table 2 presents the empirical relationship between a weaker preference for being honest (low ILC) and respondents' likelihood to work in the German public sector as a civil servant (focusing on those respondents who are currently employed). Among more dishonest respondents (i.e., those with low intrinsic lying costs), five percent work as a civil servant, whereas the

⁴⁶For further information on the methodology of the SOEP-IS see its latest report (Zweck and Rathje 2021). Note that our current analyses does not include survey weights. We will update the respective analyses when the survey weights become available.

⁴⁷The only variable without overlap was the question regarding honesty as a parenting goal for one's children (2021 SOEP edgoal5) which we consequently cannot analyze.

⁴⁸According to the 2023 Corruption Perceptions Index (CPI), Denmark is considered the least corrupt country whereas whereas India is found on rank 93.

probability substantially increases (by 4 percentage points, see Column (1)) when a participant is classified as having high ILC. When adding the pre-registered set of controls, gender, household income, age, years of education, the state a respondent lives in, and the order in which the survey items were presented (see Column (2)), the relationship between high ILC and working as a civil servant weakens. The estimated marginal effect is 3 percentage points, but becomes statistically insignificant.⁴⁹

In addition to selection into working environments in which one's preference for being (dis)honest may be beneficial, it is plausible to assume that individuals with high SIC may wish to avoid working in industries or occupations considered as immoral by others. For example, Schneider et al. (2020) study explicitly whether individuals' aversion to immoral behavior (measured in a sender receiver game without deniability) impacts their labor market outcomes and find that immoral types state a greater willingness to work in firms and industries others perceive as immoral. Following their results, we preregistered to study whether a preference for being seen as honest alters the likelihood of working in industries that could be perceived as immoral. Unfortunately, and in contrast to their laboratory and online study, the industry variables available in the German SOEP are defined at a higher level (NACE-2).⁵⁰ Consequently, we were eventually unable to directly map respondents to

Table 2: Public sector, job change, and working in immoral industries

	Civil Servant		Immoral Industry		Job Change	
	(1)	(2)	(3)	(4)	(5)	(6)
high ILC	0.04** (0.02)	0.02 (0.02)			-0.07*** (0.02)	-0.04* (0.02)
high SIC			0.01 (0.04)	0.00 (0.04)	-0.03 (0.04)	-0.02 (0.03)
high ILC and high SIC					0.03 (0.05)	0.04 (0.05)
N	1306	1294	1312	1312	1254	1254
Pseudo/ Adj. R-Squared	0.008	0.138	0.000	0.023	0.004	0.210
Controls		✓		✓		✓

Notes: The dependent variable in Columns (1) and (2) is a binary indicator for being a civil servant. The dependent variable in Columns (3) and (4) is a binary indicator for working in an industry in the bottom tercile of the ChatCPT-based morality rating. The dependent variable in Columns (5) and (6) is the fraction how often a respondent indicated a change in employment (relative to all responses regarding their employment status). Specifications (1) to (4) report marginal effects from Probit regressions. Columns (5) and (6) report coefficients from OLS regressions. Standard errors (in parentheses) are clustered at the household level. Controls include a dummy for women, household income, age, years of education, state dummies, and an indicator for the order of IPT questions. Asterisks indicate that the estimate is statistically significant at the 1% ***, 5% **, and 10% * levels.

⁴⁹Note that education strongly correlates with high ILC and also with the probability of working as a civil servant. In our specification, the variance inflation factor for education is high (22.35), indicating that including education as a control (as preregistered) may cause a potential problem of multicollinearity. As such, our results in this specification should be interpreted with a grain of salt.

⁵⁰For example, at the NACE-2 level, there is only one clearly immoral category, the Gambling and betting industry, but less than one percent of respondents actually work in this industry. Other industries typically considered as immoral (such as weapon or tobacco production) are subsumed in categories such as "Manufacture of basic metals" and "Crop and animal production, hunting and related service activities".

the industry categories provided by Schneider et al. (2020). To still shed some light on the relationship of SIC and the perceived morality of different industries, we provide an exploratory approach in Columns (3) and (4). Based on ChatGPT 4's capability to meaningfully annotate text data (Celebi and Penczynski 2024), we prompted ChatGPT (4o) to provide morality ratings on the NACE-2 industry level and estimated whether working in a particular immoral industry (i.e., in the bottom-tercile of the ratings) is related to participants' SIC.⁵¹ As shown in columns (2) and (3), we do not find a statistically significant relationship between SIC and working in a particular immoral industry.

Finally, we suspected that individuals with low intrinsic costs of lying (compared to those with high intrinsic costs) are more likely to change jobs, particularly when they have weak preferences for being seen as honest. This may be due to their reduced concern for the ethical implications of breaking commitments, deceiving employers, or navigating job negotiations with less honesty, or because they are more prone to overstating qualifications (which only yields short-term benefits). We thus preregistered to test whether job-changes are more likely to occur, when individuals have low as compared to high ILC. Columns (5) and (6) of Table 2 present coefficients from OLS regression results on the fraction of job changes reported by our SOEP respondents. Indeed, we find that participants with high ILC are less likely to change jobs than participants with low ILC (and SIC). Focusing on respondents classified as having low ILC and low SIC, the fraction of times they responded to the question whether they changed their job (in the year preceding the interview) amounts to 30 percent. As shown in Column (5), this fraction decreases on average by about 7 percentage points. Adding our set of control variables reduces this effect to four percentage points while the coefficient remains statistically significant at the ten percent level.

5.4.2 Households: Characteristics and Dynamics

Since the seminal work by G. S. Becker (1973), assortative matching has been considered an important aspect of household formation. Empirically, it has been found that partners are often similar to each other, for example in terms of socio-economic status, educational attainment, psychological characteristics or physical attributes and that such similarity is not strongly increasing in the length of the relationship (Abramitzky et al. 2011; A. Banerjee et al. 2013; Blossfeld and Timm 2003; Eika et al. 2019; Little et al. 2006; Silventoinen et al. 2003; Stevens et al. 1990; Tognetti et al. 2014). Introducing our survey items into the SOEP-IS allows us to study whether partners living in the same household are also similar in terms of individual preferences for being or being seen as honest. Accordingly, we preregistered to test whether the observed fraction of honesty preference type matches within

⁵¹We instructed ChatGPT to predict each industry's morality rating that would result from a survey of 2000 SOEP participants who have to rate the morality of different industries based on a 5-point Likert scale.

a household differs from the proportions of type matches that would result from random matching of preference types existing in the data.

We observe 444 households with a household head and a corresponding partner. Our preference measure allows us to analyze within household similarity in terms of the four preference types defined above (low ILC and low SIC, high ILC and low SIC, low ILC and high SIC, as well as high ILC and high SIC) and separately based on ILC or SIC only. To form correct expectations about the fraction of type matches with random partners, we simulate 10,000 random partner pairings for each household. That is, we repeatedly assign each of the 444 household heads to a randomly selected partner from our sample (with replacement). Based on these 10,000 random matches, we then calculate for each household head the average share of preference type mismatches. This simulation exercise allows us to estimate how many out of the 444 households we should expect to have a partner of a different preference type. This fraction amounts to 55 percent (or 244 out of 444 households). In contrast, the actual frequency of type mismatches amounts to only 203 (45.72 percent). Thus, we observe that partners are more similar than under random matching (binomial test, $p < 0.001$). Repeating this approach and focusing solely on ILC, we find that 188 out of 444 household heads with a partner (42.34 percent) are expected to be matched with a different ILC type under random matching, whereas the actual frequency of type mismatches is substantially lower (156 out of 444 households or 35.14 percent, binomial test, $p = 0.002$). Finally, focusing on SIC, we find that 22.07 percent (98 out of 444 household heads with a partner) are matched with a different SIC type, whereas with random matching, we would expect 28.15 percent (125 out of 444 households, binomial test, $p = 0.002$). Hence, also when considering intrinsic lying costs and social image costs separately, we observe results in line with assortative matching. Notably, we find that the correlations between the length of the relationships and type matches (based on preference type, ILC, and SIC) are small and statistically insignificant (Spearman's $\rho = 0.082$, $p = 0.206$; $\rho = 0.071$, $p = 0.276$; $\rho = 0.059$, $p = 0.361$). This indicates that similarity of partners may rather result from assortative matching than adjustments in moral values within households over time.

In addition to studying assortative matching, we explore whether low ILC may also result in a higher frequency of changing partners, in particular, when social image costs are low. For example, it may be less appealing to live with a partner who lies to others and does not care about how others perceive their dishonest behavior. Hence, we preregistered to study whether individuals with low ILC are more likely to change their partners. Table 3, illustrates how ILC and SIC relate to the years in current relationships (Columns 1 and 2) and the number of separations respondents reported (Columns 3 and 4). The explanatory variables are indicators for whether the respondent has high ILC, high SIC, or both, such that our baseline are respondents with low ILC and SIC. Indeed, without our set of controls, respondents with high ILC, that is, those who care about being honest, have spent more time in their current relationship. However, this coefficient becomes substantially smaller

Table 3: Relationships

	Years in current relationship		Number of separations	
	(1)	(2)	(3)	(4)
high ILC	3.46*** (1.34)	1.01 (0.88)	-0.03 (0.08)	-0.06 (0.08)
high SIC	1.93 (2.14)	-0.28 (1.36)	-0.03 (0.13)	0.00 (0.12)
high ILC and high SIC	-0.39 (3.08)	-2.39 (2.09)	-0.11 (0.18)	-0.07 (0.18)
N	1030	1030	581	581
Adj. R-Squared	0.005	0.597	-0.003	0.036
Controls		✓		✓

Notes: The dependent variables are the length of the current relationship (Columns 1-2) and the number of separations (Columns 3-4). Displayed coefficients are from OLS regressions. Standard errors clustered at the household level are in parentheses. Controls include a dummy for women, household income, age, years of education, state dummies, and an indicator for the order of IPT questions. Asterisks indicate that the estimate is statistically significant at the 1% ***, 5% **, and 10% * levels.

and statistically insignificant, when adding our set of control variables. Further, we do not find a statistically significant relationship for the number of separations; see Columns (3) and (4). These findings thus align with our earlier assortative matching interpretation.⁵²

5.4.3 The Impact of (Early) Life Experiences

Finally, we aimed to study whether (early) life experiences shape individual preferences for truth-telling (see also Abeler et al. 2024). To do so, we preregistered to focus in particular on the religiosity of parents and the exposure to different political systems. In the context of religiosity, researchers argue that parents prioritize teaching moral values to their children. Parents often act as role models, especially when their children observe them (see also Houser et al. 2016; Sutter et al. 2019) and parental behavior may influence whether children follow rules or act dishonestly (Hays and Carver 2014). Further, previous work has found a positive link between religion and moral judgments (Kirchmaier et al. 2018). We thus hypothesized that parental religiosity impacts respondents' ILC and SIC.⁵³ In Table 4, we report marginal effects of Probit models estimating whether a respondent has high ILC or high SIC depending on whether their parents are religious. We compare respondents with non-religious parents to those whose parents are either both religious or where only one parent, either the mother or father, is religious. In Columns (1) and (2), we observe a weak relationship

⁵²Note that we further preregistered to test whether individuals with children and high ILC are more likely to indicate that the honesty of their child is an important parenting goal (than individuals with low ILC), but, as noted above, there is no overlap of our SOEP-IS respondents with respondents of the 2011 SOEP wave, in which the parenting goal question was asked.

⁵³Originally, we planned to condition this analysis on the religiosity of the parent with which the individual spent their first 15 years with. However, eventually we had to rely on a general question regarding the religiosity of fathers and mothers (2020 SOEP-IS, Q437).

between parental religiosity and ILC. However, Columns (3) and (4) show that respondents with both parents being religious are more likely to exhibit high SIC, even after including our set of control variables. Although plausible, we advise caution in interpreting these results, as the data on parental religiosity is limited to a sub-sample of fewer than 300 respondents, with only five reporting that only their father is religious (none of whom exhibit high SIC).

Another important aspect that may affect individual preferences for being or being seen as honest is the political system that respondents have experienced. In the German context, it has been shown that experiencing the GDR regime may shape individuals' preferences for state-provided services but also affect their social trust (Alesina and Fuchs-Schündeln 2007; Rainer and Siedler 2009). The GDR was an autocratic state that maintained control through one of the most extensive surveillance networks in history, relying on ordinary citizens as informants to secretly gather information within their social circles, which resulted in a profound and lasting erosion of interpersonal trust (Lichter et al. 2021). The state's influence began early through the education system, which emphasized ideological conformity and the importance of outwardly conforming to rules and social norms. We thus hypothesized that having lived in the GDR or having experienced GDR schooling may affect individual preferences for being (ILC), and particularly, for being seen as honest (SIC).

Table 5 regresses the indicator variable for having high ILC (Columns (1) to (3)) or high SIC (Columns (4) to (6)) on a variable that indicates whether a respondent lived in the GDR before the German reunification in 1989 and Table 6 regresses the indicator for having high ILC (Columns (1) to (3)) or high SIC (Columns (4) to (6)) on a variable that indicates whether the respondent attended

Table 4: Religiosity of Parents, ILC, and SIC

	High ILC		High SIC	
	(1)	(2)	(3)	(4)
both religious	-0.08 (0.26)	-0.21 (0.23)	1.13*** (0.18)	1.04*** (0.18)
father religious	0.08 (0.22)	0.13 (0.19)	-0.99*** (0.12)	-0.92*** (0.13)
mother religious	-0.01 (0.14)	-0.06 (0.14)	-0.08 (0.14)	0.01 (0.13)
N	299	297	297	283
Pseudo R-Squared	0.000	0.087	0.015	0.071
Controls		✓		✓

Notes: The dependent variable is an indicator for having high ILC (Columns 1 and 2) or high SIC (Columns 3 and 4). Displayed coefficients are marginal effects from probit regressions. Standard errors clustered at the household level are in parentheses. Controls include a dummy for women, household income, age, years of education, state dummies, and an indicator for the order of IPT questions. Asterisks indicate that the estimate is statistically significant at the 1% ***, 5% **, and 10% * levels.

Table 5: Lived in GDR, ILC, and SIC

	High ILC			High SIC		
	(1)	(2)	(3)	(4)	(5)	(6)
Lived in GDR	-0.04 (0.03)	-0.01 (0.03)	-0.03 (0.03)	0.04* (0.02)	0.01 (0.03)	0.00 (0.03)
N	2233	2233	2233	2231	2231	2231
Pseudo R-Squared	0.001	0.002	0.023	0.002	0.003	0.037
Former GDR State		✓	✓		✓	✓
Controls (w/o state FE)			✓			✓

Notes: The dependent variable is an indicator for having high ILC or high SIC. Lived in GDR indicates that the respondent lived in the GDR before 1989. We report marginal effects from Probit regressions in all columns. Clustered standard errors (at the household level) in parentheses. Controls include a dummy for women, household income, age, years of education, and an indicator for the order of IPT questions. Asterisks indicate that the estimate is statistically significant at the 1% ***, 5% **, and 10% * levels.

Table 6: Attended school last in GDR, ILC, and SIC

	High ILC			High SIC		
	(1)	(2)	(3)	(4)	(5)	(6)
School in GDR	-0.02 (0.03)	0.02 (0.03)	-0.04 (0.03)	0.07*** (0.02)	0.07** (0.03)	0.05* (0.03)
N	2233	2233	2233	2231	2231	2231
Pseudo R-Squared	0.000	0.002	0.023	0.006	0.006	0.039
Former GDR State		✓	✓		✓	✓
Controls (w/o state FE)			✓			✓

Notes: The dependent variable is an indicator for having high ILC or high SIC. School in GDR indicates that the school the respondent attended last was in the GDR. We report marginal effects from Probit regressions in all columns. Clustered standard errors (at the household level) in parentheses. Controls include a dummy for women, household income, age, years of education, and an indicator for the order of IPT questions. Asterisks indicate that the estimate is statistically significant at the 1% ***, 5% **, and 10% * levels.

school last in the GDR. Because having lived in the GDR or attending school in the GDR may also proxy living in a former GDR state today, we add a dummy variable for whether the respondent lives in a former GDR state today (instead of state fixed effects) in Columns (2) and (5). Finally, in Columns (3) and (6), we include our additional control variables.

As shown in Table 5, we find no statistically significant relationship between living in the GDR and ILC (Columns (1) to (3)), and if at all a weak relationship between living in the GDR and SIC when adding the dummy variable for whether the respondent lives in a former GDR state today and our controls (Columns (5) to (6)). However, we do find that having been in school last in the GDR – and thus having experienced the regime at a younger age – relates statistically significantly to higher SIC (Table 6, columns (4) to (6)). The marginal probability of having high SIC is about 7 percentage points larger, even when controlling for whether a respondent still lives in a state that formerly belonged to the GDR (Column (5)) and about 5 percentage points higher when controlling for

our additional set of control variables (Column (6)). Hence, our data allows for novel insights in terms of the malleability of honesty preferences. They are in line with the idea that early life experiences may shaped the preference for being seen as honest but do not indicate a strong relationship between such experiences and participants intrinsic preference for being honest.

6 Replication exercise

To establish the validity and robustness of empirical findings, replications are crucial. Apart from large-scale replication approaches (Camerer et al. 2016a,b), direct or so-called ‘self-replications’ are an appealing measure to establish robustness of results within and across different participant pools (Englmaier et al. 2024; List 2003, 2004a,b, 2006; Schultze et al. 2019; Shah et al. 2019). Following this idea, we conduct a replication exercise by administering our incentivized IPT measure, the mind game, and the survey measure in an additional sample representative of the UK in terms of gender, age, and ethnicity.⁵⁴ This allows us to assess i) the heterogeneity of preferences within a representative sample and the robustness of type distributions across samples, ii) whether the incentivized measure and the survey measure remain predictive for behavior of a more diverse sample, and iii) which socio-demographic characteristics relate systematically to intrinsic preferences for being and being seen as honest.

6.1 Type distribution

We use our IPT paradigm and classify participants based on their incentivized willingness to pay for information (low ILC if $WTP_{info} > 0$) and their incentivized willingness to pay for the favorable switch (high SIC if $WTP_{switch} > 0$). Table 7 shows the resulting type distribution. Again, we identify substantial heterogeneity in IPT and a negative correlation between intrinsic lying costs and social image costs (Spearman’s $\rho = -0.289$, $p\text{-value} < 0.001$). To study how much heterogeneity exists in the type distribution across samples, we additionally compare the data from all samples in which we administered the incentivized IPT measure. As shown in Figure 10, we find fairly robust type distributions across all three samples. The most prevalent type is the one with low ILC and high SIC, i.e., intrinsically willing to lie but concerned about their social image (between 35% and 39% of the samples). The least frequent type is the one that cares about both motives, i.e., high ILC and high SIC (between 7% and 11%). In the representative and the student samples, the second most frequent type cares about neither motive (31% and 28%, respectively), whereas this type is the third most frequent in in the convenience sample (26%). In sum, we document substantial heterogeneity in preferences for truth-telling across individuals (in all three samples), with a substantial fraction of participants caring about only one of the

⁵⁴We use Prolific’s feature to recruit the representative sample ($n = 500$). Although this sample encompasses older participant who are less familiar with studies run on Prolific, similarly many participants needed to be excluded according to our pre-registered exclusion criteria as in the Prolific convenience sample (eight vs. eleven percent). These participants were replaced by new invites adhering to the original quotas.

Table 7: IPT Types

	Low SIC	High SIC	
Low ILC	31.0 %	39.0 %	70.0 %
High ILC	22.0 %	8.0 %	30.0 %
	53.0 %	47.0 %	

Notes: Representative sample (n=500). A participant is classified as low ILC if their WTP_{info} is strictly larger than zero, and as high ILC otherwise. Conversely, a participant is classified as low SIC if their WTP_{switch} is zero, and as high SIC otherwise.

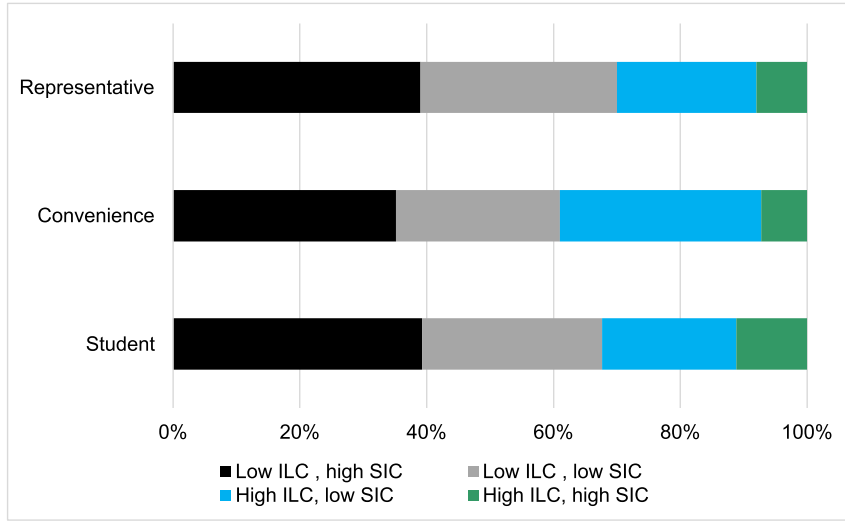


Figure 10: Comparison of the type distribution across samples

two underlying motivations (or none). Additional ANOVA analyses show that the between-sample variation of the WTP_{info} and the WTP_{switch} is substantially lower than the within-sample variation (Bartlett's equal-variances tests, both p-values < 0.001). This further illustrates that individual-level heterogeneity in preferences for truth-telling is more prevalent than heterogeneity across samples.

6.2 Predictiveness of the incentivized and the survey measure

As pre-registered, we also replicate the assessment of our measures' predictiveness using the mind game, both for the incentivized measure (analogously to Section 4.2.1) and the survey measure (analogously to Section 5.2). Figure 11 illustrates payoffs from the mind game for the four preference types. The top-panel shows participants with low SIC, among which we clearly observe that participants with high ILC reported substantially lower payoffs than participants with low ILC. A similar picture arises when comparing participants in the bottom panels (i.e. those with high SIC). These observations are also reflected in a significant negative correlation between intrinsic lying costs, captured by WTP_{info} , and the claimed payoff in the mind game (Spearman's $\rho = 0.205$, $p < 0.0001$).

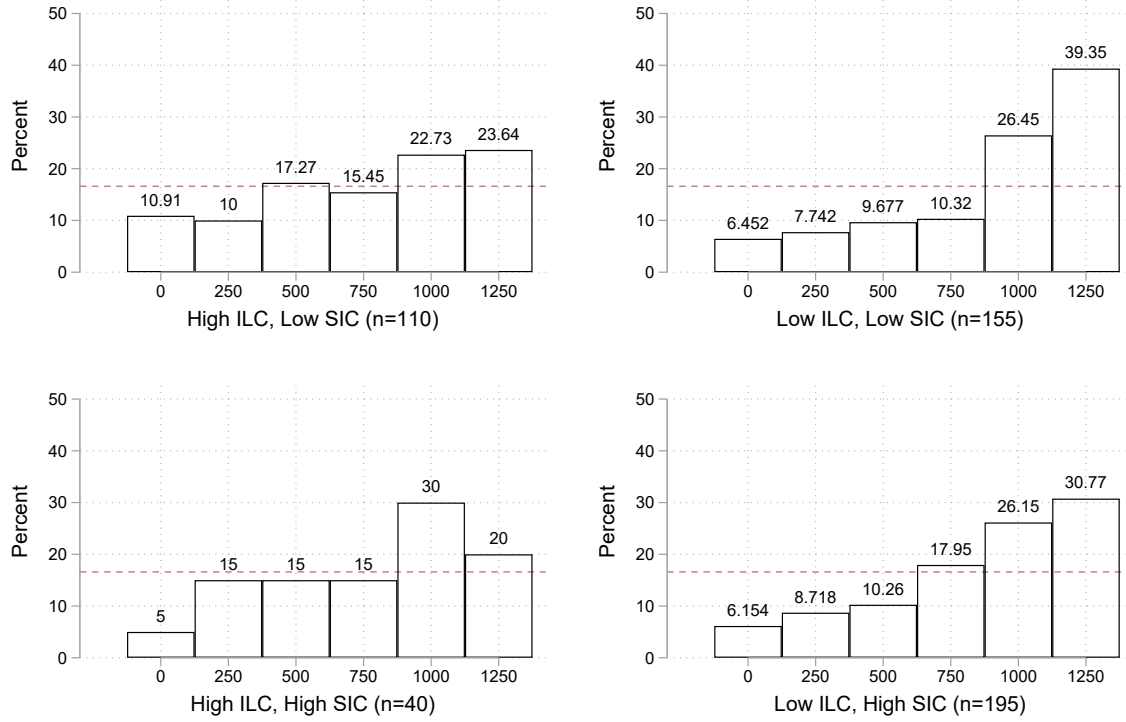


Figure 11: Claimed payoffs in the mind game by IPT types

Notes: The figure shows the distribution of claimed payoffs in the mind game by types (based on incentivized measure) in the representative sample ($n=500$). *ILC* denotes intrinsic lying costs and *SIC* denotes social image costs. The dashed horizontal line represents the expected fraction of reports for each payoff when all participants report honestly.

Although visual inspection further indicates a tendency of participants' with high SIC to report the highest payoff less often (comparison of top vs. bottom panels), we do not find a statistically significant correlation between receiving the highest payoff and WTP_{switch} (Spearman's $\rho = -0.013$, $p=0.771$), even when focusing exclusively on individuals with low intrinsic lying costs (Spearman's $\rho = -0.060$, $p=0.265$). However, focusing on participants that are generally willing to lie (i.e., those with low ILC), we find that the fraction receiving the highest payoff is significantly lower with high as compared to low SIC (33.3 vs. 47.2, χ^2 -test, $p=0.093$, see also right panels in Figure 11).

Concerning the survey module, we proceed as in Section 5.2 along our pre-registration. Most importantly, we replicate that the survey question corresponding to ILC meaningfully captures variation in the incentivized experimental measure, both at the intensive (i.e., WTP_{info}) and the extensive margin (the ILC type classification).⁵⁵ In terms of social image costs, we find a somewhat weaker relationship

⁵⁵In particular, WTP_{info} is smaller for those who report zero tails compared to those who report more than zero tails (for a true outcome of zero, MWU, $p<0.001$). Regarding the extensive margin, the fraction of participants reporting zero tails (for a true outcome of zero) differs between participants with high ILC as compared to low ILC (χ^2 -test, $p<0.0001$).

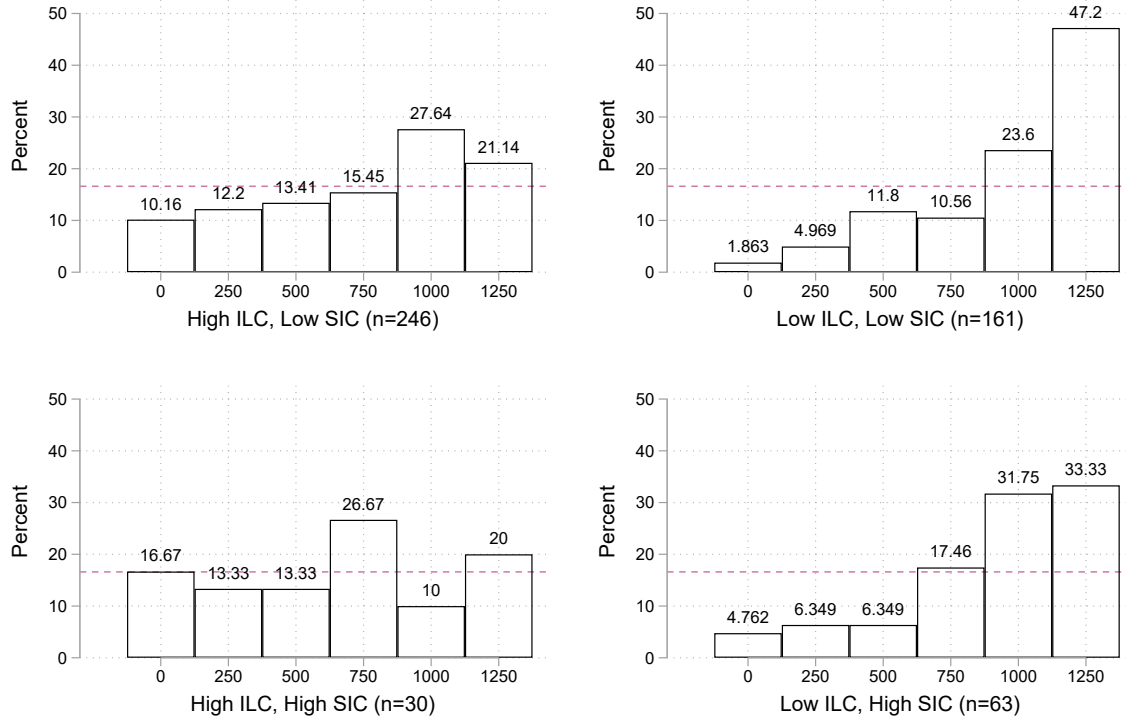


Figure 12: Claimed payoffs in the mind game by types (based on survey measure)

Notes: The figure shows the distribution of claimed payoffs in the mind game by types (defined based on the survey measure) in the representative sample ($n=500$). *ILC* denotes intrinsic lying costs and *SIC* denotes social image costs. The dashed horizontal line represents the expected fraction of reports for each payoff when all participants report honestly.

that is only statistically significant at the extensive margin.⁵⁶ Nonetheless, the survey questions are predictive of behavior in the mind game (see Figure 12 and Appendix C for the corresponding tests).

6.3 Correlates of ILC and SIC

The representative sample allows us also to shed light on the relationship of personal characteristics and individual preferences for truth-telling. Previous correlational evidence from aggregate lying measures (such as the die-rolling paradigm, Fischbacher and Föllmi-Heusi 2013) indicate that women tend to lie less than men (see, e.g., the meta-studies by Abeler et al. (2019), Capraro (2018), and Gerlach et al. (2019)),⁵⁷ and further, a (somewhat weaker) positive relationship of age and honesty. For example, the meta-study by Gerlach et al. (2019) finds less lying for older individuals. Abeler et al. (2019) report a similar tendency, albeit in their data, the positive relationship of age and honesty is not

⁵⁶The WTP_{switch} is larger for those reporting less than 4 than for those who report 4 (for a true outcome of 4) (MWU, $p=0.0415$). However, the fraction of decision makers reporting less than 4 tails (for a true outcome of 4) is not different among DMs with high SIC and low SIC (χ^2 -test, $p=0.147$).

⁵⁷Single studies investigating this relationship either find that women lie less (e.g., Conrads et al. 2013; Dreber and Johannesson 2008; Erat and Gneezy 2012; Fischbacher and Föllmi-Heusi 2013; Friesen and Gangadharan 2012; Grosch and Rau 2017; Kajackaite and Gneezy 2017; Kocher et al. 2018) or the same (e.g., Abeler et al. 2014; Belot and Schröder 2013; Childs 2012; Ezquerro et al. 2018; Gravert 2013; Lundquist et al. 2009; Muehlheusser et al. 2015; Pate 2018).

Table 8: Correlates of WTP_{info} and WTP_{switch}

	Representative Sample		Convenience Sample		Student Sample	
	WTP Info	WTP Switch	WTP Info	WTP Switch	WTP Info	WTP Switch
Female	-35.23** (17.56)	-11.29 (10.82)	-52.80*** (17.83)	-6.64 (10.97)	-49.12 (33.16)	2.61 (29.43)
Age	-0.46 (0.53)	0.37 (0.33)	-2.83*** (0.57)	-0.61* (0.34)	13.63 (18.56)	13.70 (18.39)
Political preference	1.68 (6.89)	4.65 (4.18)	3.76 (6.26)	9.86** (3.86)	-12.74 (12.42)	-14.68 (10.06)
N	498	498	450	450	323	324
R-Squared	.01	.0098	.059	.019	.011	.007

Notes: Data from the representative sample (Columns 1 and 2), the convenience sample (Columns 3 and 4), and the student sample (Columns 5 and 6). DV is either WTP_{info} or WTP_{switch} , as indicated in the column headings. OLS regressions and robust standard errors in parentheses. Female is a binary variable, age is measured in years in the representative and the convenience sample and in six brackets in the student sample. Political preference is measured on a scale from 1 (left) to 7 (right). Asterisks indicate that the estimate is statistically significant at the 1% ***, 5% **, and 10% * levels.

robust to different empirical specifications.⁵⁸ In contrast to our approach, this literature did not aim at disentangling intrinsic lying costs and social image costs. For example, not reporting the maximum payoff a reporting paradigm is consistent with both high intrinsic lying costs and high social image costs (Abeler et al. 2019; Fischbacher and Föllmi-Heusi 2013). Our measure instead enables us to correlate gender and age separately with intrinsic lying costs and social image costs (across three samples). As we additionally elicited political preferences in a consistent way in all three samples, we also explore the relationship of individual preferences for truth-telling and political orientation.⁵⁹

Table 8 displays the results. In line with previous findings, we identify the most robust relationship for gender. In particular, being female is negatively related to the WTP_{info} , i.e., women are less willing to lie than men due to an intrinsic preference for honesty. While the coefficient is negative in all three samples, the size and statistical significance vary. Interestingly, we see no significant relationship between gender and the preference for being seen as honest. Thus, the well-documented relationship of more honest women is likely driven by gender difference in intrinsic preferences for truth-telling rather than higher social image costs among women.

Akin to previous findings in the literature, our results regarding age are mixed. While the negative correlation with WTP_{info} in the convenience sample suggests older individuals to be more honest, the coefficient in the representative sample is smaller and statistically insignificant.⁶⁰ As for

⁵⁸Single studies find that older individuals lie less (e.g., Conrads et al. 2013; Friesen and Gangadharan 2013; Glätzle-Rützler and Lergetporer 2015) or the same (e.g., Abeler et al. 2014; Bucciol and Piovesan 2011; Conrads and Lotz 2015).

⁵⁹Few studies report correlations of lying with political orientation. One exception is Abeler et al. (2014), who find no strong relationship in reported payoffs and political preferences. Concerning other characteristics that may relate to IPT, we elicited religiosity, but only in the representative sample. We find that religiosity correlates negatively with WTP_{info} , but not statistically significantly with WTP_{switch} .

⁶⁰In the student sample, the coefficient is even positive, however, that majority of students is between 21 and 25 years old, such that there is a lack of variation in the explanatory

gender, we see no consistent relationship between age and social image costs. Finally, we find no consistent relationship between political preferences and preferences for being or being seen as honest.

7 Conclusion

This paper provides a comprehensive assessment of individual preferences for truth-telling. Various models of lying costs propose the distinction between intrinsic costs of lying and social image costs of lying (Abeler et al. 2019; Gneezy et al. 2018; Khalmetski and Sliwka 2019). So far, these types of costs could only be measured on an aggregate level. We propose a novel method of assessing these two types of costs on an individual level and independently of each other, but in one coherent setting. Our experimental measure of individual preferences for truth-telling elicits two willingnesses to pay that are indicative of the individual's preference for being honest and for being seen as honest. Using this measure in three different samples, we find substantial heterogeneity in preference types. Further, we show that the preference types captured by our experimental measure are predictive of behavior in two other incentivized experimental paradigms which are commonly used to study honesty and deception: a mind game (Greene and Paxton 2009; Jiang 2013; Kajackaite and Gneezy 2017). and a sender-receiver game (Gneezy 2005; Gneezy et al. 2013). We also present evidence that social image costs identified with our incentivized IPT measure meaningfully relate to lying for social image in a paradigm (without any material gains) in which individuals trade off social image and honesty.

While our incentivized experimental measure offers detailed insights into the relation of ILC and SIC and allows for more fine-grained classifications of preference types, it may not be practical to elicit preferences when research time is scarce. For this reason, we also developed a 2-min survey module that consists of only two questions. The module allows researchers to proxy individual preferences for truth-telling due to intrinsic lying costs and social image costs. In a within-individual comparison of classifications, we document meaningful overlap between classified types based on the experimental and the survey measure, and highlight the survey measures predictiveness for the mind game. Thus, the survey measure can serve as a reasonable alternative when the incentivized procedure is not applicable or when a binary type classification is sufficient.

Our results from the incentivized measure and the survey measure underscore the critical role of both intrinsic lying costs and preferences for being seen as honest. First, the two measures reveal systematic heterogeneity in preference types, and the incentivized measure highlights the remarkable stability of preference type distributions across different samples (extending recent findings on the stability of honesty preferences, see, e.g., Bortolotti et al. 2022). Second, by independently measuring the two preference components, our approach offers new and valuable insights into potential determinants of (un)ethical behavior. On the one hand, our measure allows for a better understanding

of social image concerns and observability (see, e.g., Fries et al. 2021; Köbis et al. 2016; Van de Ven and Villeval 2015; Villeval 2024) as well as the potential consequences of varying communication channels, dynamics, and interaction partners (see, e.g., Cohn et al. 2022; Leib et al. 2024; Rilke et al. 2021). On the other hand, independently measuring both preference components reveals that early life experiences may particularly shape preferences for being seen as honest rather than the intrinsic preference for being honest (Abeler et al. 2024).

Finally, our approach opens several promising avenues for future research. For example, it appears promising to administer our measures across other representative samples or panels, and linking these with administrative data. This will further enhance our understanding of fraudulent real world behaviors and allows for exploring the cultural underpinnings of individual preferences for being honest and the desire to be perceived as honest. As we observe systematic heterogeneity in preference types but stable type distributions across samples, future research may also investigate whether the effectiveness of interventions aimed at reducing dishonest behavior—such as moral appeals, norm nudges, and honesty oaths— depends not only on the informational content or framing (see, e.g., Jacquemet et al. 2019; Kingsuwankul et al. 2023; Zickfeld et al. 2024) but also on the underlying type distribution within a population. Hence, our novel measurement may contribute to a better understanding of institutional effectiveness in addressing fraudulent behaviors as well as to the design of more informed and targeted interventions.

References

- Abeler, Johannes, Anke Becker, and Armin Falk (2014): "Representative evidence on lying costs." *Journal of Public Economics* 113, pp. 96–104.
- Abeler, Johannes, Armin Falk, and Fabian Kosse (2024): "Malleability of preferences for honesty." *The Economic Journal*, ueae044.
- Abeler, Johannes, Daniele Nosenzo, and Collin Raymond (2019): "Preferences for Truth-Telling." *Econometrica* 87 (4), pp. 1115–1153.
- Abramitzky, Ran, Adeline Delavande, and Luis Vasconcelos (2011): "Marrying up: the role of sex ratio in assortative matching." *American Economic Journal: Applied Economics* 3 (3), pp. 124–157.
- Akerlof, George A. and Rachel E. Kranton (2000): "Economics and Identity." *The Quarterly Journal of Economics* 115 (3), pp. 715–753.
- Albertazzi, Andrea (2021): "Individual cheating in the lab: a new measure and external validity." *Theory and Decision*.
- Alesina, Alberto and Nicola Fuchs-Schündeln (2007): "Good-bye Lenin (or not?): The effect of communism on people's preferences." *American Economic Review* 97 (4), pp. 1507–1528.
- Balafoutas, Loukas, Adrian Beck, Rudolf Kerschbamer, and Matthias Sutter (2013): "What drives taxi drivers? A field experiment on fraud in a market for credence goods." *Review of Economic Studies* 80 (3), pp. 876–891.
- Balafoutas, Loukas, Simon Czermak, Marc Eulerich, and Helena Fornwagner (2020): "Incentives for Dishonesty: an Experimental Study With Internal Auditors." *Economic Inquiry* 58 (2), pp. 764–779.
- Banerjee, Abhijit, Esther Duflo, Maitreesh Ghatak, and Jeanne Lafortune (2013): "Marry for what? Caste and mate selection in modern India." *American Economic Journal: Microeconomics* 5 (2), pp. 33–72.
- Banerjee, Ritwik, Tushi Baul, and Tanya Rosenblat (2015): "On self selection of the corrupt into the public sector." *Economics Letters* 127, pp. 43–46.
- Barfort, Sebastian, Nikolaj A Harmon, Frederik Hjorth, and Asmus Leth Olsen (2019): "Sustaining honesty in public service: The role of selection." *American Economic Journal: Economic Policy* 11 (4), pp. 96–123.
- Barron, Kai, Agne Kajackaite, and Silvia Saccardo (2022): "Image Concerns and Lying Behavior." Available at SSRN: <https://ssrn.com/abstract=4111941>.
- Bašić, Zvonimir and Simone Quercia (2022): "The Influence of Self and Social Image Concerns on Lying." *Games and Economic Behavior*.
- Becker, Gary S (1973): "A theory of marriage: Part I." *Journal of Political economy* 81 (4), pp. 813–846.
- Becker, Gordon M., Morris H. Degroot, and Jacob Marschak (1964): "Measuring Utility by a Single-Response Sequential Method." *Behavioral Science* 9, pp. 226–232.
- Belot, Michèle and Marina Schröder (2013): "Sloppy work, lies and theft: A novel experimental design to study counterproductive behaviour." *Journal of Economic Behavior and Organization* 93, pp. 233–238.
- Blossfeld, Hans-Peter and Andreas Timm (2003): *Who marries whom?: educational systems as marriage markets in modern societies*. Vol. 12. Springer Science & Business Media.
- Bortolotti, Stefania, Felix Kölle, and Lukas Wenner (2022): "On the persistence of dishonesty." *Journal of Economic Behavior & Organization* 200, pp. 1053–1065.
- Bott, Kristina M, Alexander W Cappelen, Erik Ø Sørensen, and Bertil Tungodden (2020): "You've got mail: A randomized field experiment on tax evasion." *Management Science* 66 (7), pp. 2801–2819.
- Buccioli, Alessandro and Marco Piovesan (2011): "Luck or cheating? A field experiment on honesty with children." *Journal of Economic Psychology* 32 (1), pp. 73–78.
- Camerer, Colin et al. (2016a): "Evaluating replicability of laboratory experiments in economics." *Science* 351 (6280), pp. 1433–1436.

- Camerer, Colin et al. (2016b): "Evaluating replicability of laboratory experiments in economics." *Science* 351 (6280), pp. 1433–1436.
- Capraro, Valerio (2018): "Gender differences in lying in sender-receiver games: A meta-analysis." *Judgment and Decision Making* 13 (4), pp. 345–355.
- Celebi, Can and Stefan Penczynski (2024): *Using Large Language Models for Text Classification in Experimental Economics*. Tech. rep. School of Economics, University of East Anglia, Norwich, UK.
- Charness, Gary, Celia Blanco-Jimenez, Lara Ezquerra, and Ismael Rodriguez-Lara (2019): "Cheating, incentives, and money manipulation." *Experimental Economics* 22, pp. 155–177.
- Chen, Daniel L., Martin Schonger, and Chris Wickens (2016): "oTree—An open-source platform for laboratory, online, and field experiments." *Journal of Behavioral and Experimental Finance* 9, pp. 88–97.
- Childs, Jason (2012): "Gender differences in lying." *Economics Letters* 114 (2), pp. 147–149.
- Cohn, Alain, Tobias Gesche, and Michel André Maréchal (2022): "Honesty in the digital age." *Management Science* 68 (2), pp. 827–845.
- Cohn, Alain, Michel André Maréchal, David Tannenbaum, and Christian Lukas Zünd (2019): "Civic honesty around the globe." *Science* 365 (6448), pp. 70–73.
- Conrads, Julian, Bernd Irlenbusch, Rainer Michael Rilke, and Gari Walkowitz (2013): "Lying and team incentives." *Journal of Economic Psychology* 34, pp. 1–7.
- Conrads, Julian and Sebastian Lotz (2015): "The effect of communication channels on dishonest behavior." *Journal of Behavioral and Experimental Economics* 58, pp. 88–93.
- Dohmen, Thomas, Armin Falk, David Huffman, Uwe Sunde, Jürgen Schupp, and Gert G. Wagner (2011): "Individual risk attitudes: Measurement, determinants, and behavioral consequences." *Journal of the European Economic Association* 9 (3), pp. 522–550.
- Dreber, Anna and Magnus Johannesson (2008): "Gender differences in deception." *Economics Letters* 99 (1), pp. 197–199.
- Dufwenberg, Martin and Martin A Dufwenberg (2018): "Lies in disguise—A theoretical analysis of cheating." *Journal of Economic Theory* 175, pp. 248–264.
- Eika, Lasse, Magne Mogstad, and Basit Zafar (2019): "Educational assortative mating and household income inequality." *Journal of Political Economy* 127 (6), pp. 2795–2835.
- Eliasz, Kfir and Andrew Schotter (2010): "Paying for confidence: An experimental study of the demand for non-instrumental information." *Games and Economic Behavior* 70 (2), pp. 304–324.
- Englmaier, Florian, Stefan Grimm, Dominik Grothe, David Schindler, and Simeon Schudy (2024): "The Effect of Incentives in Nonroutine Analytical Team Tasks." *Journal of Political Economy* 132 (8), pp. 2531–2880.
- Erat, Sanjiv and Uri Gneezy (2012): "White lies." *Management Science* 58 (4), pp. 723–733.
- Ezquerra, Lara, Georgi I. Kolev, and Ismael Rodriguez-Lara (2018): "Gender differences in cheating: Loss vs. gain framing." *Economics Letters* 163, pp. 46–49.
- Feess, Eberhard, Peter J Jost, and Anna Ressi (2024): "Fake News and the Problem of Disregarding True Messages: Theory and Experimental Evidence." *Available at SSRN* 4810742.
- Feess, Eberhard, Florian Kerzenmacher, and Gerd Muehlheusser (2023): "Morally questionable decisions by groups: Guilt sharing and its underlying motives." *Games and Economic Behavior* 140, pp. 380–400.
- Feess, Eberhard, Florian Kerzenmacher, and Yuriy Timofeyev (2022): "Utilitarian or deontological models of moral behavior—What predicts morally questionable decisions?" *European Economic Review* 149, p. 104264.
- Fehrler, Sebastian, Urs Fischbacher, and Maik Schneider (2020a): "Honesty and Self-Selection into Cheap Talk." *Economic Journal* 130 (632), pp. 2468–2496.
- Fehrler, Sebastian, Urs Fischbacher, and Maik T Schneider (2020b): "Honesty and Self-Selection into Cheap Talk." *The Economic Journal* 130 (632), pp. 2468–2496.

- Fischbacher, Urs and Franziska Föllmi-Heusi (2013): "Lies in disguise-an experimental study on cheating." *Journal of the European Economic Association* 11 (3), pp. 525–547.
- Fries, Tilman, Uri Gneezy, Agne Kajackaite, and Daniel Parra (2021): "Observability and lying." *Journal of Economic Behavior & Organization* 189, pp. 132–149.
- Friesen, Lana and Lata Gangadharan (2012): "Individual level evidence of dishonesty and the gender effect." *Economics Letters* 117 (3), pp. 624–626.
- (2013): "Designing self-reporting regimes to encourage truth telling: An experimental study." *Journal of Economic Behavior & Organization* 94, pp. 90–102.
- Geraldes, Diogo, Franziska Heinicke, and Duk Gyoo Kim (2022): "The Effect of Chosen or Given Luck on Honesty." *CESifo Working Paper*.
- Gerlach, Philipp, Kinneret Teodorescu, and Ralph Hertwig (2019): "The Truth About Lies: A Meta-Analysis on Dishonest Behavior." *Psychological Bulletin* 145 (1), pp. 1–44.
- Glätzle-Rützler, Daniela and Philipp Lergetporer (2015): "Lying and age: An experimental study." *Journal of Economic Psychology* 46, pp. 12–25.
- Gneezy, Uri (2005): "Deception: The role of consequences." *American Economic Review* 95 (1), pp. 384–394.
- Gneezy, Uri, Agne Kajackaite, and Joel Sobel (2018): "Lying Aversion and the Size of the Lie." *American Economic Review* 108 (2), pp. 419–453.
- Gneezy, Uri, Bettina Rockenbach, and Marta Serra-Garcia (2013): "Measuring lying aversion." *Journal of Economic Behavior and Organization* 93, pp. 293–300.
- Gravert, Christina (2013): "How luck and performance affect stealing." *Journal of Economic Behavior and Organization* 93, pp. 301–304.
- Greene, Joshua D and Joseph M Paxton (2009): "Patterns of neural activity associated with honest and dishonest moral decisions." *Proceedings of the National Academy of Sciences* 106 (30), pp. 12506–12511.
- Grolleau, Gilles, Martin G Kocher, and Angela Sutan (2016): "Cheating and loss aversion: Do people cheat more to avoid a loss?" *Management Science* 62 (12), pp. 3428–3438.
- Grosch, Kerstin and Holger A Rau (2017): "Gender differences in honesty: The role of social value orientation." *Journal of Economic Psychology* 62, pp. 258–267.
- Hanna, Rema and Shing-Yi Wang (2017): "Dishonesty and selection into public service: Evidence from India." *American Economic Journal: Economic Policy* 9 (3), pp. 262–90.
- Hays, Chelsea and Leslie J Carver (2014): "Follow the liar: the effects of adult lies on children's honesty." *Developmental Science* 17 (6), pp. 977–983.
- Houser, Daniel, John A List, Marco Piovesan, Anya Samek, and Joachim Winter (2016): "Dishonesty: From parents to children." *European Economic Review* 82, pp. 242–254.
- Hugh-Jones, David (2016): "Honesty, beliefs about honesty, and economic growth in 15 countries." *Journal of Economic Behavior & Organization* 127, pp. 99–114.
- Jacquemet, Nicolas, Stéphane Luchini, Julie Rosaz, and Jason F Shogren (2019): "Truth telling under oath." *Management Science* 65 (1), pp. 426–438.
- Jiang, Ting (2013): "Cheating in mind games: The subtlety of rules matters." *Journal of Economic Behavior and Organization* 93, pp. 328–336.
- Kajackaite, Agne and Uri Gneezy (2017): "Incentives and cheating." *Games and Economic Behavior* 102, pp. 433–444.
- Kashdan, Todd B., Matthew W. Gallagher, Paul J. Silvia, Beate P. Winterstein, William E. Breen, Daniel Terhar, and Michael F. Steger (2009): "The curiosity and exploration inventory-II: Development, factor structure, and psychometrics." *Journal of Research in Personality* 43 (6), pp. 987–998.
- Kerschbamer, Rudolf and Matthias Sutter (2017): "The Economics of Credence Goods—a Survey of Recent Lab and Field Experiments." *CESifo Economic Studies* 63 (1), pp. 1–23.
- Khalmetski, Kiryl and Dirk Sliwka (2019): "Disguising lies - Image concerns and partial lying in

- cheating games." *American Economic Journal: Microeconomics*, pp. 1–38.
- Kingsuwanikul, Sorrravich, Chloe Tergiman, and Marie Claire Villeval (2023): "Why do oaths work? Image concerns and credibility in promise keeping." *Image Concerns and Credibility in Promise Keeping* (September 16, 2023).
- Kirchmaier, Isadora, Jens Prüfer, and Stefan T Trautmann (2018): "Religion, moral attitudes and economic behavior." *Journal of Economic Behavior & Organization* 148, pp. 282–300.
- Köbis, NC, JW Van Prooijen, F Righetti, and PAM Van Lange (2016): "The look over the shoulder– Corruption and cheating decreases in the presence of another person." *Manuscript in Preparation*.
- Kocher, Martin G., Simeon Schudy, and Lisa Spantig (2018): "I lie? We lie! Why? Experimental evidence on a dishonesty shift in groups." *Management Science* 64 (9), pp. 3995–4008.
- Konrad, Kai A., Tim Lohse, and Salmai Qari (2014): "Deception choice and self-selection - The importance of being earnest." *Journal of Economic Behavior and Organization* 107 (PA), pp. 25–39.
- Lang, Matthias and Simeon Schudy (2023): "(Dis) honesty and the value of transparency for campaign promises." *European Economic Review* 159, p. 104560.
- Lefebvre, Mathieu, Pierre Pestieau, Arno Riedl, and Marie Claire Villeval (2015): "Tax evasion and social information: an experiment in Belgium, France, and the Netherlands." *International Tax and Public Finance* 22, pp. 401–425.
- Leib, Margarita, Nils Köbis, Rainer Michael Rilke, Marloes Hagens, and Bernd Irlenbusch (2024): "Corrupted by algorithms? how ai-generated and human-written advice shape (dis) honesty." *The Economic Journal* 134 (658), pp. 766–784.
- Lichter, Andreas, Max Löffler, and Sebastian Sieglöcher (2021): "The long-term costs of government surveillance: Insights from Stasi spying in East Germany." *Journal of the European Economic Association* 19 (2), pp. 741–789.
- List, John A (2003): "Does market experience eliminate market anomalies?" *Quarterly Journal of Economics* 118 (1), pp. 41–71.
- (2004a): "Neoclassical theory versus prospect theory: Evidence from the marketplace." *Econometrica* 72 (2), pp. 615–625.
- (2004b): "The nature and extent of discrimination in the marketplace: Evidence from the field." *Quarterly Journal of Economics* 119 (1), pp. 49–89.
- (2006): "The behavioralist meets the market: Measuring social preferences and reputation effects in actual transactions." *Journal of Political Economy* 114 (1), pp. 1–37.
- Little, Anthony C, D Michael Burt, and David I Perrett (2006): "Assortative mating for perceived facial personality traits." *Personality and Individual Differences* 40 (5), pp. 973–984.
- Lundquist, Tore, Tore Ellingsen, Erik Gribbe, and Magnus Johannesson (2009): "The aversion to lying." *Journal of Economic Behavior and Organization* 70 (1-2), pp. 81–92.
- Mazar, Nina, On Amir, and Dan Ariely (2008): "The dishonesty of honest people: A theory of self-concept maintenance." *Journal of Marketing Research* 45, pp. 633–644.
- Muehlheusser, Gerd, Andreas Roider, and Niklas Wallmeier (2015): "Gender differences in honesty: Groups versus individuals." *Economics Letters* 128, pp. 25–29.
- Murphy, Ryan O, Kurt A Ackermann, and Michel JJ Handgraaf (2011): "Measuring social value orientation." *Judgment and Decision Making* 6 (8), pp. 771–781.
- Pate, Jennifer (2018): "Temptation and cheating behavior: Experimental evidence." *Journal of Economic Psychology* 67 (May), pp. 135–148.
- Rainer, Helmut and Thomas Siedler (2009): "Does democracy foster trust?" *Journal of Comparative Economics* 37 (2), pp. 251–269.
- Rilke, Rainer Michael, Anastasia Danilov, Ori Weisel, Shaul Shalvi, and Bernd Irlenbusch (2021): "When leading by example leads to less corrupt collaboration." *Journal of Economic Behavior & Organization* 188, pp. 288–306.
- Saccardo, Silvia and Marta Serra-Garcia (2023): "Enabling or limiting cognitive flexibility? evidence

- of demand for moral commitment." *American Economic Review* 113 (2), pp. 396–429.
- Schneider, Florian, Fanny Brun, and Roberto A Weber (2020): "Sorting and wage premiums in immoral work." *University of Zurich, Department of Economics, Working Paper* (353).
- Schultze, Thomas, Juergen Huber, Michael Kirchler, and Andreas Mojzisch (2019): "Replications in economic psychology and behavioral economics." *Journal of Economic Psychology* 75 (5), p. 102199.
- Shah, Anuj K, Sendhil Mullainathan, and Eldar Shafir (2019): "An exercise in self-replication: Replicating Shah, Mullainathan, and Shafir (2012)." *Journal of Economic Psychology* 75, p. 102127.
- Shalvi, Shaul, Jason Dana, Michel J.J. Handgraaf, and Carsten K.W. De Dreu (2011): "Justified ethicality: Observing desired counterfactuals modifies ethical perceptions and behavior." *Organizational Behavior and Human Decision Processes* 115, pp. 181–190.
- Shalvi, Shaul, Ori Eldar, and Yoella Bereby-Meyer (2012): "Honesty requires time (and lack of justifications)." *Psychological science* 23 (10), pp. 1264–1270.
- Silventoinen, Karri, Jaakko Kaprio, Eero Lahelma, Richard J Viken, and Richard J Rose (2003): "Assortative mating by body height and BMI: Finnish twins and their spouses." *American Journal of Human Biology* 15 (5), pp. 620–627.
- Stevens, Gillian, Dawn Owens, and Eric C Schaefer (1990): "Education and attractiveness in marriage choices." *Social Psychology Quarterly*, pp. 62–70.
- Sutter, Matthias (2009): "Deception Through Telling the Truth?! Experimental Evidence from Individuals and Teams." *The Economic Journal* 119 (534), pp. 47–60.
- Sutter, Matthias, Claudia Zoller, and Daniela Glätzle-Rützler (2019): "Economic behavior of children and adolescents—A first survey of experimental economics results." *European Economic Review* 111, pp. 98–121.
- Tognetti, Arnaud, Claire Berticat, Michel Raymond, and Charlotte Faurie (2014): "Assortative mating based on cooperativeness and generosity." *Journal of evolutionary biology* 27 (5), pp. 975–981.
- Trocinska, Matylda (2020): "Is dishonesty contagious? in relation to culture and national identity." *Master Thesis*.
- Van de Ven, Jeroen and Marie Claire Villeval (2015): "Dishonesty under scrutiny." *Journal of the Economic Science Association* 1, pp. 86–99.
- Verschuere, Bruno et al. (2018): "Registered replication report on Mazar, Amir, and Ariely (2008)." *Advances in Methods and Practices in Psychological Science* 1 (3), pp. 299–317.
- Villeval, Marie Claire (2024): "The social determinants of unethical behavior." *Research Handbook on Unethical Behavior*.
- Weisel, Ori and Shaul Shalvi (2015): "The collaborative roots of corruption." *Proceedings of the National Academy of Sciences* 112 (34), pp. 10651–10656.
- Zickfeld, Janis H et al. (2024): "Effectiveness of ex ante honesty oaths in reducing dishonesty depends on content." *Nature Human Behaviour*, pp. 1–19.
- Zweck, Bettina and Martin Rathje (2021): *SOEP-IS 2020 – Survey Report on the 2020 SOEP Innovation Sample*. Tech. rep. 986. Berlin: DIW/SOEP.

Online Appendix

A Additional results: Student sample (n=331)

A.1 Heterogeneity in IPT

We briefly present heterogeneity in preferences for truth-telling for the student sample. Figure A.1 displays similar heterogeneity as in our general population sample (Figure 2). The figure plots the distributions (WTP_{info} in Panel A.1a) and (WTP_{switch} in Panel A.1b). Table A.1 shows the resulting type classification.

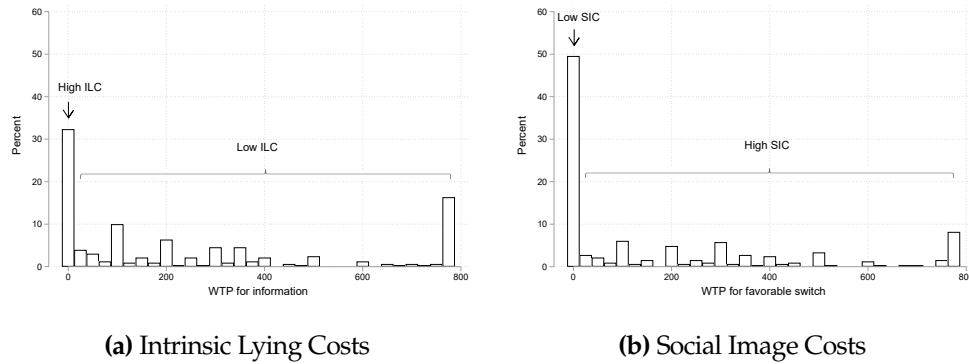


Figure A.1: Distributions of ILC and SIC

Notes: Panel a presents the distribution of WTP_{info} , where smaller values imply larger ILC. Panel b presents the distribution of WTP_{switch} , where smaller values imply smaller SIC. Student sample (n=331).

Table A.1: IPT Types

	Low SIC	High SIC	
Low ILC	28.4 %	39.3 %	67.7 %
High ILC	21.2 %	11.2 %	32.4 %
	49.6 %	50.5 %	

Notes: Student sample (n=331). A participant is classified as high intrinsic lying costs (ILC) if their WTP_{info} is equal to zero, and as low ILC otherwise. Conversely, a participant is classified as high social image cost (SIC) if their WTP_{switch} is strictly larger than zero, and as low SIC otherwise.

A.2 Reports

The focus of our IPT measure is on the intention to be honest and the intention to be perceived as honest. Therefore, we do not discuss participants' reports in the main text. For completeness, however, we briefly present results on reporting behavior in the convenience sample below, even

Table A.2: Reports in high- and low-stakes IPT tasks

	Report low in low-stakes task	Report high in low-stakes task	
Report low in high-stakes task	44.1 %	12.7 %	56.8 %
Report high in high-stakes task	26.6 %	16.6 %	43.2 %
	70.7 %	29.3 %	

Notes: Student sample (n=331).

though this behavior may be confounded by participants' reevaluation of their earlier choices, choice inconsistencies, or reporting decisions aimed at avoiding costs associated with a favorable switch.

Honest reporting would predict the high-paying item to be reported with a probability of 12.5% in each task. Table A.2 shows evidence of misreporting at the aggregate level, as the fractions of high-paying item reports are substantially larger for both the high-stakes task (43.2 %) as well as the low-stakes task (29.3 %). Interestingly, misreporting rates are higher in the high-stakes task despite the high-paying items being known to everyone in the low-stakes task. Of those who buy the information, 67 % report the high-paying item. This implies that 33 % report the low-paying item after learning the high-paying outcome. Part of this appears to be driven by social image concerns as not everyone who was willing to pay for the favorable switch was willing to pay the randomly determined price for it.

A.3 Robustness: Curiosity and Risk

Our measure of the intention to be honest (WTP_{info}) is based on the assumption that the reason for wanting to learn the high-paying item is that participants intend to (are willing to) lie. Additional factors that could have influenced whether participants wanted to learn the correspondence of items to payoffs before reporting are curiosity (preferences for non-instrumental information Eliaz and Schotter 2010) and –conditional on being willing to misreport– participants' risk preferences. To control for these factors, we implemented an additional task in the student sample that mirrored the situation participants faced when choosing their WTP_{info} , but isolated decisions regarding information and risk preferences.

To this end, participants were informed that they would participate in a lottery in which they selected one out of eight pieces of a pie. If this piece coincided with the piece the computer randomly selected with equal probability, they would earn a payoff of 1000 ECUs, whereas if they selected one of the other seven pieces, they would earn a payoff of 200 ECUs (akin to the randomly generated outcome in the IPT task). Importantly, in the lottery task, participants had no opportunity to behave dishonestly. Still, we elicited whether they were willing to pay for (non-instrumental) information regarding the lottery. To do so, we informed participants that they would be randomly assigned one

of eight lotteries (with identical prospects), which only differed in terms of the color of the pie they would select the item from (blue, green, yellow, purple, pink, turquoise, red, or olive).

Mirroring the situation participants faced in WTP_{info} , we asked for their WTP for receiving information that they would otherwise receive right after their decision. That is, we elicited participants' WTP to learn which color (e.g., blue) their lottery had before they learned the outcome of the lottery they participated in ($WTP_{curiosity}$). Because the color of the lottery did not affect the lottery's prospects in any way, we truly captured a willingness to pay for non-instrumental information.

To control for risk preferences, we informed participants, after they stated their $WTP_{curiosity}$, that they could also avoid playing the lottery at all and instead secure a certain payoff of 1000 ECUs. To enable them to do so, we elicited participants' willingness to pay (between 25 and 775 ECUs) to ensure the safe payoff of 1000 ECUs instead of participating in the lottery (WTP_{risk}). Participants were informed that the possible costs they would incur to learn the color of the lottery would only become payoff-relevant if they actually participated in the lottery. This sequence of choices allowed us to elicit preferences for information even for those who wished to incur costs to ensure a safe payment. Complementing the incentivized measures, we additionally elicited participants' curiosity (Kashdan et al. 2009)⁶¹ and risk (Dohmen et al. 2011) preferences using non-incentivized survey questions.

To evaluate the importance of curiosity and risk preferences, we correlate our incentivized and non-incentivized measures with the WTP_{info} from the IPT task and use them as additional controls when evaluating the predictiveness of the IPT measure in the mind game. Notably, we find a weak and marginally statistically significant correlation between WTP_{info} and $WTP_{curiosity}$ (Spearman's $\rho = 0.092$, $p = 0.094$) and no correlation between WTP_{info} and the non-incentivized curiosity measure (Spearman's $\rho = -0.018$, $p = 0.747$). We thus conclude that curiosity does not play a crucial role in the decision to reveal the high-paying item in IPT.

Turning to risk preferences, we find a significant correlation between WTP_{info} and WTP_{risk} (Spearman's $\rho = 0.327$, $p < 0.001$), but not between WTP_{info} and the general risk question (Spearman's $\rho = 0.081$, $p = 0.144$). As such, the intensive margin of our the ILC measure may be affected by risk preferences.⁶² Importantly, further analyses reveal that risk (and curiosity) preferences do not confound the explanatory power of our IPT measure when predicting behavior in the mind game. Table A.3 displays regressions analyzing the determinants of the payoffs participants received in the mind game. Columns (1) shows the predictive power of WTP_{info} without controlling for curiosity and risk preferences. In Column (2) we add WTP_{risk} and $WTP_{curiosity}$ as control variables. Comparing the coefficients for WTP_{info} in (1) and (2) reveals that measures for risk and curiosity impact neither the direction nor the size or significance of the coefficient of WTP_{info} . For completeness, we repeat this

⁶¹For curiosity, we ask all ten items used by Kashdan et al. (2009) and form a simple average for each person.

⁶²Note again that risk preferences can only play a role for participants who are willing to lie, such that or type classification based on ($WTP_{info} = 0$) cannot be affected by risk preferences.

analysis for WTP_{switch} in Columns (3) and (4). In Column (5), we add both WTP_{info} and WTP_{switch} , and in Column (6), we further add the interaction between the two. The influence of WTP_{info} on the payoff in the mind game remains significant and positive in both specifications and is also robust to including WTP_{risk} and $WTP_{\text{curiosity}}$ (Column (7)) as well as to including additional control variables (gender, income, age, political orientation, Column (8)). Column (9) includes the IPT type classifications instead of the WTPs, and Column (10) adds WTP_{risk} and $WTP_{\text{curiosity}}$. These analyses underline the robustness of the predictive power of our IPT measure.

Table A.3: Correlates of Payoff in Mind Game

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
WTP _{info}	0.188*** (0.0625)	0.187*** (0.0682)			0.197*** (0.0627)	0.234*** (0.0742)	0.234*** (0.0809)	0.226*** (0.0834)		
WTP _{risk}		-0.0155 (0.0679)		0.0542 (0.0674)			-0.0131 (0.0703)	0.000958 (0.0738)		0.0188 (0.0676)
WTP _{curiosity}		0.0475 (0.0816)		0.0836 (0.0831)			0.0517 (0.0826)	0.0332 (0.0867)		0.0672 (0.0800)
WTP _{switch}			-0.0121 (0.0706)	-0.0352 (0.0731)	-0.0536 (0.0702)	0.0105 (0.106)	0.0121 (0.111)	0.0146 (0.110)		
WTP _{info} × WTP _{switch}						-0.000185 (0.000198)	-0.000191 (0.000202)	-0.000216 (0.000208)		
Low ILC, low SIC									194.2*** (54.41)	188.1*** (56.41)
High ILC, high SIC									34.36 (72.38)	29.03 (74.33)
Low ILC, high SIC									72.25 (56.30)	64.34 (60.10)
Constant	904.6*** (25.86)	903.7*** (33.97)	953.8*** (22.70)	925.8*** (33.81)	912.0*** (27.79)	901.9*** (31.11)	899.6*** (36.29)	790.2*** (92.21)	864.3*** (47.31)	853.1*** (49.48)
Observations	331	331	331	331	331	331	331	323	331	331
Further controls	no	no	no	no	no	no	no	yes	no	no

Standard errors in parentheses. Student sample, n=331.

Further controls: gender (female dummy); income; age (categories from 1: <21 to 6: 51-60); politics (from 1: left to 7: right)

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

A.4 Observers

We use the strategy method to elicit observers' rating, i.e. observers rate the character both of a person who reports a likely, low-paying outcome and of a person who reports an unlikely, high-paying outcome. We encourage observers to take the rating task seriously by incentivizing the ratings on the two scorecards as a coordination game. One of the two scorecards is then randomly chosen to be payoff-relevant. We compare each of the four ratings on this scorecard to the ones of a randomly matched observer and pay 200 ECU for each rating that is the same. At the very end of the experiment, observers are shown the Player-IDs of all players whose report resulted in an unlikely, high-paying outcome as well as the IDs of the players whose report resulted in a likely, low-paying outcome.⁶³

We examine how observers actually rated DMs' reports that resulted in a high and unlikely or low and likely payoff. Figure A.2 displays how observers rate DMs' honesty based on the information that DM's report resulted in a high, unlikely payoff (left panel) or in a low, likely payoff (right panel). Clearly, observers rate DMs more critically if they reported the unlikely item which yields a high payoff. The average honesty rating equals 44.8, with many observers stating that the probability of DM being an honest person is 50 percent or only 25 percent. For the more likely item which yields a low payoff, the picture changes. The most frequently chosen options are 75 percent and 100 percent and the average rating amounts to 71.88 percent. Given that we used the strategy method and asked each observer for their rating for both scenarios, we can also look at the within-subject variation across the two outcomes. We find that 75 percent of observers rate DMs more negatively if DM's report results in an unlikely high payoff (Wilcoxon signed-rank test, $p=0.002$). Thus, the social image of a DM did in fact suffer if they reported an unlikely, high-paying item.

⁶³They are further shown i) the two scorecards that observers generated themselves for unlikely, high-paying outcomes and likely-low paying outcomes, ii) the two scorecards the matched observer generated and iii) the two scorecards that result from averaging own and partner's scores.

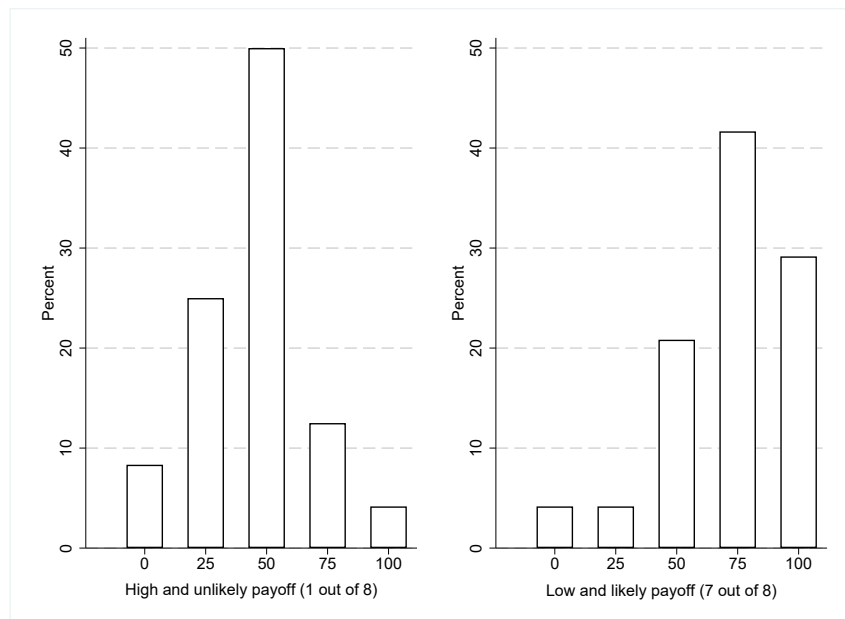


Figure A.2: Observer ratings

Notes: The figure shows how often a certain honesty rating was chosen by observers. The left panel shows ratings for the high and unlikely payoff, whereas the right panel shows ratings for the low and likely payoff. Student sample (n=24).

B Additional results: Prolific convenience sample (general population, n=471)

B.1 The mind game

To assess the robustness of our IPT measure's predictiveness, we included the mind game also in the Prolific convenience sample. In this section, we briefly replicate our analyses performed earlier in the student sample. Figure B.3 shows payoffs in the mind game by types in the convenience sample. Again, we observe a clear shift in payoffs from high to low ILC types. Those classified as having low ILC in our IPT task again claim significantly higher payoffs in the mind game than those with high ILC (844.3 vs 649.7, MWU, $p < 0.001$), and are more likely to claim the highest payoff (χ^2 -test, $p < 0.001$). Consistent with our main results and the corresponding hypothesis, WTP_{info} and the payoff in the mind game are positively correlated (Spearman's $\rho = 0.242$, $p < 0.001$).

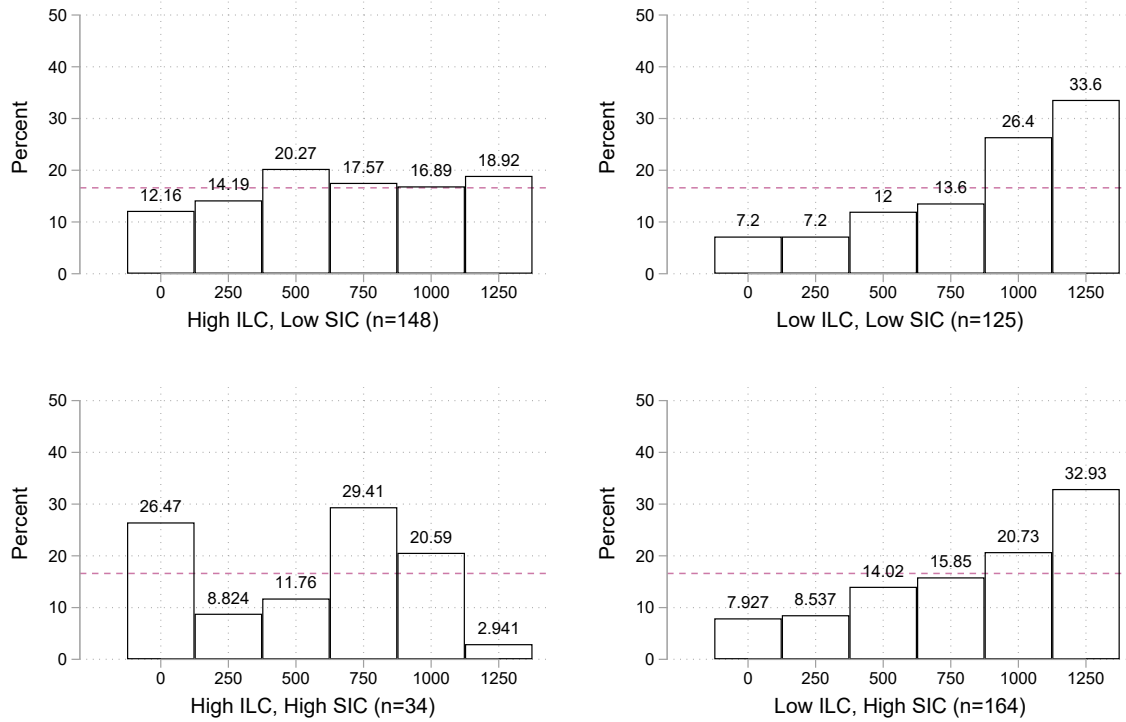


Figure B.3: Claimed Payoffs in mind game by IPT types

Notes: The figure shows the distribution of claimed payoffs in the mind game by IPT types. General population sample (n=471). The red horizontal line represents the expected fraction of reports for each payoff when all participants report honestly.

B.2 Reports

Similar to the additional analyses in the student sample, we briefly present reporting behavior results for the convenience sample below for completeness. As reporting behavior may be influenced by participants' reevaluation of earlier choices, inconsistencies in decision-making, or attempts to avoid costs associated with a favorable switch, we refrain from drawing broader conclusions based on reporting behavior.

Table B.4 displays the fractions of reports of the high- and low-paying item in the high- and low-payoff tasks. Honest reporting would predict the high-paying item to be reported with a probability of 12.5% in each task. The table shows evidence of misreporting at the aggregate level, as the fractions reporting high-paying items are substantially larger in both tasks than expected under honesty. Despite the high-paying item being known to all in the low-stakes task, the fraction of reports resulting in the high payoff are larger in the high-stakes task than in the low-stakes task (33.97% vs 24.20%). Of those who buy the information, 55 % report the high-paying item. This implies that 45 % report the low-paying item after learning the high-paying outcome. Part of this appears to be driven by social image concerns as not all participants who were willing to pay for the favorable switch were willing to pay the randomly determined price. Notably, and reassuringly those participants who bought both the information and the favorable switch, everyone reported the high-paying items in the high-stakes task.

Table B.4: Reports in high- and low-stakes IPT tasks

	Report low in low-stakes task	Report high in low-stakes task	
Report low in high-stakes task	53.08 %	12.95 %	66.03 %
Report high in high-stakes task	22.72 %	11.25 %	33.97 %
	75.80 %	24.20 %	

Notes: General population sample (n=471).

C Additional results: Prolific Representative Sample (n=500)

C.1 Reports

For completeness, we also briefly present results on reporting behavior in the representative sample below. Recall that honest reporting would predict the high-paying item to be reported with a probability of 12.5% in each task whereas, again, we find evidence for misreporting at the aggregate level (see Table C.5). The fractions of high-paying item reports are substantially larger than 12.5% for both the low-stakes task (27.6 %) as well as the high-stakes task (34.8 %). As in the other samples, misreporting rates are higher in the high-stakes task despite the high-paying items being known to everyone in the low-stakes task. Of those who buy the information, 58 % report the high-paying item. This implies that 42 % report the low-paying item after learning the high-paying outcome. Again, this appears to be at least partially driven by social image concerns as not all participants who were willing to pay for the favorable switch but received a relatively high (randomly determined) price. Notably, again, those participants who bought both the information and the favorable switch all reported the high-paying items in the high-stakes task.

Table C.5: Reports in high- and low-stakes IPT tasks

	Report low in low-stakes task	Report high in low-stakes task	
Report low in high-stakes task	53.2 %	12 %	65.02 %
Report high in high-stakes task	19.2 %	15.6 %	34.8 %
	72.4 %	27.6 %	

Notes: Representative sample (n=500).

C.2 Explanations of behavior

To better understand participants' motivation behind their choices in the IPT task, we asked them four questions regarding the favorable switch and four questions regarding the information on a ten point Likert-scale in the post-experimental questionnaire as follows: "Below we list factors that could have influenced your decision to pay for the information. To what extent do you agree with the following statements".

- I considered the [switch/information] useful for my report
- I selected the wrong button
- I cared about how the researchers would perceive my decision to buy the [switch/information]
- I wanted to save money by not buying the [switch/information]

Below, we present the results separately by participants' ILC / SIC type. Table C.6 shows mean agreement to the statements for individuals with high ILC (Column 1) and individuals with low ILC (Column 2), as well as the difference in means. While answers may suffer from ex-post rationalization, they are in line with our interpretation of the measure: Individuals trade off the costs and the usefulness of information, and those who decide to buy (i.e., those with low ILC) are more likely to state that information is useful and less likely to indicate that they wanted to save money. Further, we find that errors appear rare (clicking the wrong button) and are slightly more often reported to occur by low ILC types (potentially as these participants use errors as an exp-post excuse for buying the information). In addition, we find that low ILC types (i.e., those wanted to buy the information for the high stake report) indicate being more concerned about how they are perceived by the researchers than high ILC types (those not buying the information). This finding suggests that social desirability did not generally keep participants willing to lie from buying the information. Finally, we find that both low and high ILC types indicated a reasonable level of understanding of the instructions (5.4-5.5 on a 7 point Likert scale, where 7 indicates "very high"). As the most pronounced differences between the two groups arise for the two items that capture the core trade-off between the costs and the usefulness of information, we are confident that our measure captures ILC well.

Table C.7 presents results regarding participants' decision for (or against) the favorable switch. Akin to the findings regarding ILC, we find that those who opt for the option at hand (here the favorable switch) are more likely to report that the option (the switch) was useful and less likely to state that they wanted to save money with their choice. Also, errors regarding the switch decision are rarely reported, but if so, slightly more often among those who opt for the switch. Importantly, we again find that those opting for the option at hand (the switch) are the ones who are more likely

Table C.6: ILC type and explanation of behavior regarding information

Variable	(1) Mean ILC low	(2) Mean ILC high	(3) Difference
Considered info useful	7.034 (2.515)	3.953 (2.970)	-3.081*** (0.278)
Clicked wrong button	1.491 (1.542)	1.120 (0.655)	-0.371*** (0.098)
Cared about researchers' perception	3.977 (2.783)	3.407 (2.922)	-0.570** (0.282)
Wanted to save money by not buying	3.897 (2.775)	6.507 (3.491)	2.610*** (0.322)
Understanding	5.391 (1.245)	5.547 (1.229)	0.155 (0.121)
Observations	350	150	500

Notes: Representative sample (n=500). Likert-scale questions that ask whether the following was important in the decision from 1 (not at all) to 10 (to a great extent). *Understanding* is self-rated understanding of the instructions (on a 7-point scale where, 7 indicates "very high").

Table C.7: SIC type and explanation of behavior regarding favorable switch

Variable	(1) Mean SIC low	(2) Mean SIC high	(3) Difference
Considered switch useful	3.419 (2.579)	5.723 (2.588)	2.305*** (0.232)
Clicked wrong button	1.279 (1.065)	1.723 (1.836)	0.444*** (0.137)
Cared about researchers' perception	3.362 (2.716)	4.843 (2.807)	1.480*** (0.248)
Wanted to save money by not buying	6.204 (3.276)	3.936 (2.790)	-2.268*** (0.272)
Understanding	5.506 (1.191)	5.362 (1.295)	-0.144 (0.112)
Observations	265	235	500

Notes: Representative sample (n=500). Likert-scale questions that ask whether the following was important in the decision from 1 (not at all) to 10 (to a great extent). *Understanding* is self-rated understanding of the instructions (on a 7-point scale where, 7 indicates "very high").

to state that they are concerned about how this choice might appear to the researchers. Hence, social desirability again did not seem to stop them from opting for the switch. Finally, also for SIC, we find a reasonable level of understanding of the instructions among both low and high SIC types. Overall, and similarly to the responses regarding ILC types, our findings underline that our measurement of SIC captures the core trade-off between the costs and benefits of appearing honest towards others.

C.3 Survey measure and mind game

As in Section 5.3, we find that lying costs proxied by the survey question are informative for claimed payoffs in the mind game: those with low ILC claim significantly larger payoffs (959 vs. 743 points; MWU, $p < 0.001$). Focusing on individuals with low SIC, we again find that claiming the maximum payoff occurs more than twice as often for those with low ILC (see upper two panels of Figure 12). Regarding social image costs, we also replicate our findings from Section 5.3. Across all participants, SIC are not predictive of the payoff claimed (MWU, $p = 0.696$) or claiming the maximum payoff (χ^2 , $p = 0.649$). Conditional on low intrinsic lying costs, participants with low SIC are more likely to claim the highest payoff (47% vs. 33%; χ^2 , $p = 0.060$), whereas SIC do not matter for those with high intrinsic lying costs (21% vs 20% claim the largest payoff; χ^2 , $p = 0.885$). These results closely mirror the results regarding the predictiveness of the survey measure in the convenience sample (Section 5.3) and provide further evidence of the validity of our survey module.

D Data collection logistics

To provide full transparency in terms of data collection, we briefly summarize the evolution of this project in this section. An overview of our preregistrations can be found in Table D.8. At the end of 2020, we initiated this project with an elaborate experimental framework aimed at examining the interplay between intrinsic lying costs (ILC) and social image concerns (SIC). In this initial design, participants could achieve varying payoff levels (low, medium, or high) within a single reporting paradigm and were additionally asked for their willingness to pay for appearing more honest to observers contingent upon the payoff they might receive. However, this design proved operationally intricate and introduced a potential confound, undermining the independent measurement of ILC and SIC. Recognizing these limitations, we opted for a fundamental redesign, prioritizing a simplified experimental setup that preserved the independence of the two constructs, which is the basis of this manuscript. We implemented this new design for the first time in July 2021, with a student sample ($n=331$). In the same year, we also ran the experiment and the survey questions in the online convenience sample. In 2023, we additionally conducted the treatments to establish internal validity and, in 2024, we replicated the experiment in the representative sample. Data from the German SOEP (including the results from our survey module) were received in May 2024.

Table D.8: Pre-Registration

Date	Sample	Short description	Registration	Link
16.12.2020	student sample	early, more complex design (not included)	#54433	https://aspredicted.org/ztq8-j2fz.pdf
13.07.2021	student sample	main experiment	#70550	https://aspredicted.org/pg8v-67t4.pdf
05.12.2021	Prolific convenience	main experiment	#82028	https://aspredicted.org/74py-sbc7.pdf
18.04.2023	Prolific convenience	0 vs 2 observer	#129232	https://aspredicted.org/nwt2-ym4v.pdf
04.05.2023	Prolific convenience	Lie vs NoLie	#131061	https://aspredicted.org/6tqc-9j6p.pdf
12.02.2024	UK representative	replication	#161756	https://aspredicted.org/kt5w-5b5x.pdf
28.03.2024	SOEP	survey module only		https://osf.io/fqsw

Finally, Table D.9 shows which tasks have been administered in which sample, and when exactly this data collection took place.

Table D.9: Data presented in this paper

Data collection	Sample	N	Experimental tasks	Survey questions
10.08.-04.11.2021	Students	331	IPT, mind game, sender-receiver, knowledge reporting, risk and curiosity task	no
06.12.-09.12.2021	Prolific convenience	471	IPT, mind game	yes
20.04.-24.04.2023	Prolific convenience	499	IPT treatments: ZeroObservers or TwoObservers	no
08.05.-10.05.2023	Prolific convenience	503	IPT treatments: LieNoObservers or NoLieNoObservers	no
04.03.-15.03.2023	UK representative	500	IPT, mind game	yes
May – Nov 2023 (received April '24)	SOEP-IS	2,387	none	yes

E Instructions and Screenshots

E.1 Experimental Measure (IPT)

Decision-Makers

The following section displays screenshots of the instructions we used on Prolific. Instructions for the student sample are available upon request.

General instructions

- This study consists of two parts and the final questionnaire. At the beginning of each part, additional instructions will appear on your screen.
- Please read the instructions carefully and answer the control questions.
- Control questions must be answered correctly before you can proceed.
- **We will pay you a bonus payment of £0.23 for the control questions. If you repeatedly fail to answer the questions correctly you will not receive the bonus payment.**
- If you fail to answer several control questions correctly on the first try, you may not receive this additional bonus payment.

Please note that this study needs to be completed on a laptop, computer, or tablet.

You will need a pen and a piece of paper. Please make sure now that you have them ready.

Next

Payment

You will only receive a monetary payoff if you complete the entire study.

You will receive

- £3.75 for the completion
- £0.23 for answering the control questions correctly
- An additional bonus payment depending on your choices in one of the two different parts.

One part of this study has been randomly determined to be payoff relevant for you. For this part, we will convert the points earned into British Pounds. The conversion rate is:

$$£1.00 = 260 \text{ points}$$

Which part is relevant for your payment will only be shown at the end of the study. As each choice may be relevant for your bonus payment, please carefully consider all your choices.

Next

Part 1

Task 1: Vegetable

We have assigned a participant number to each participant. You are participant 1.

We now ask you to spin a wheel of fortune on an external website to generate a random outcome and write the outcome down on a piece of paper. The written note is only for you to remember your outcome, as we will ask you about the outcome you saw at a later stage.

To do so,

1. Use the link provided below to open the random outcome generator website in a new browser tab or window. <https://wordwall.net/resource/26056229>
2. On the website, click on "Start" and then on "Spin It". The wheel of fortune will spin and generate a random vegetable.
3. Write down the vegetable that was randomly selected on a piece of paper.
4. When you are done, return to the study page.

Please now confirm that you have written down the randomly generated vegetable.

☐ I confirm that I have written down the vegetable.

Next

Task 2: Fruit

We now ask you to spin a different wheel of fortune to generate another random outcome and write it down on a piece of paper. As before, the written note is only for you to remember your outcome, as we will ask you about the outcome you saw at a later stage.

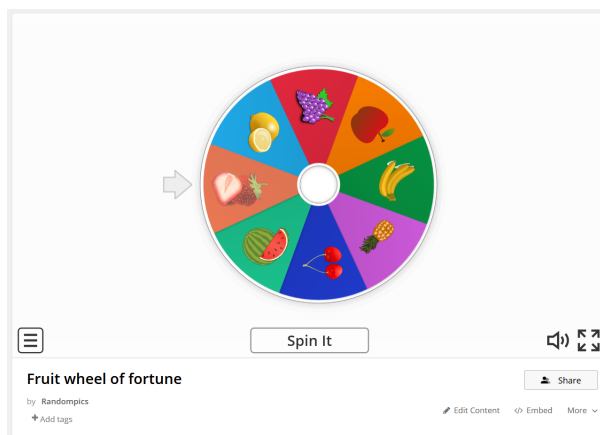
1. Use the link provided below to open the random outcome generator website in a new browser tab or window. <https://wordwall.net/resource/26056405>
2. On the website, click on "Start" and then on "Spin It". The wheel of fortune will spin and generate a random fruit.
3. Write down the fruit that was randomly selected on a piece of paper.
4. When you are done, return to the study page.

Please now confirm that you have written down the randomly generated fruit.

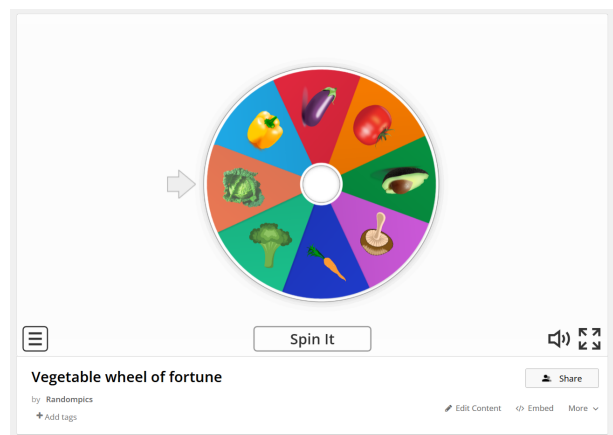
☐ I confirm that I have written down the fruit.

Next

Screenshots Wheels of Fortune for Random Draw (on external website)



(a) Wheel of Fortune - fruit



(b) Wheel of Fortune - vegetables

Reports and payments

Later, we will ask you to tell us which two items (that is, which fruit and which vegetable) you have randomly generated with the two wheels of fortune (task 1 and task 2). These two reports will determine your payoff for this study and we will sum up the points resulting from your two reports.

Payoff

For each report you make (fruit and vegetable), you may either receive a lower or a higher payoff. It is determined randomly which items yield a lower or a higher payoff.

For your **fruit report** you will receive either 10 points or 2 points:

- One fruit yields 10 points
- The seven other fruits yield 2 points

For example, if Banana has randomly been determined to yield a payoff of 10 points, you will receive...

- 10 points if you report "Banana",
- 2 points if you report any of the other fruits.

For your **vegetable report** you will receive either 1000 points or 200 points:

- One vegetable yields 1000 points
- The seven other vegetables yield 200 points

For example, if Tomato has randomly been determined to yield a payoff of 1000 points, you will receive...

- 1000 points if you report "Tomato",
- 200 points if you report any of the other vegetables.

Information on your payoff

Currently you do not know which fruit and which vegetable has been assigned to the higher payoffs, but each fruit and each vegetable has the same chance of yielding the higher payoff.

Before reporting, you will learn for one of the foods (fruit or vegetable) which item yields the higher payoff.

Further, you may (or may not) learn which item yields the higher payoff for the other food.

Next

Control Questions

To proceed, please indicate whether the following statements are true or false:

Payoffs resulting from the reported fruit and the reported vegetable will be summed up:

- ☒ True
☐ False

For both foods, it is randomly determined which item yields the higher payoff:

- ☒ True
☐ False

Before reporting, I will learn which item yields the higher payoff for one of the two reports (fruit or vegetable). I may (or may not) learn this for the other report as well.

- ☒ True
☐ False

For each food (fruit and vegetable)...

- 2 out of the 8 items (25 percent of items) are associated with the higher payoff:

- ☐ True
☒ False


- 6 out of the 8 items (75 percent of items) are associated with the lower payoff:

- ☐ True
☒ False

Next

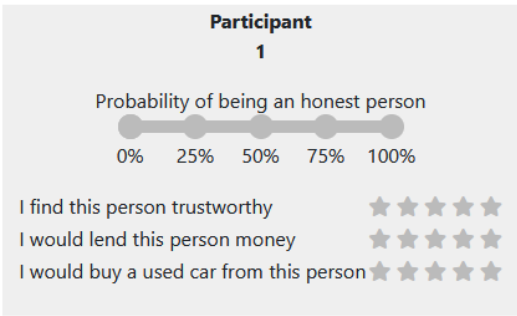
Scorecard and Observers

Nobody except you knows which fruit and vegetable were generated by the wheels of fortune. However, we will inform other prolific participants about whether you received a high or low payoff from one of your reports.



We call these other participants “Observers”. Observers do not make any reports themselves and do not affect your payoffs. Observers’ task is to reflect on what to conclude about your character based on your reporting behavior in either the fruit or the vegetable task.

For each participant, two Observers fill in a scorecard (see example below).



Information provided to Observers

The Observers will fill in your scorecard based on whether your report resulted in

- the high payoff that is unlikely (1 out of 8 items assigned)
- or the low payoff that is more likely (7 out of 8 items assigned).

NOTE:

- Observers will only learn whether your report resulted in the high or the low payoff, but are not informed about the exact size of the payoff.
- Observers will only learn about the result of one of your reports (either fruit or vegetable). The computer has randomly determined whether observers are informed about the fruit or the vegetable task.

Knowledge of Observers

What else do Observers know?

Before filling in the scorecard and reflecting on your character, Observers will learn that

- ...your task is to report one item that was randomly selected out of 8 possible items by a Wheel of Fortune,
- ...your report determines your payoff,
- ...only one item yields a high payoff and seven items yield a low payoff,
- ...you may know before reporting which item yields a high or a low payoff.

What is unknown to Observers?

Observers are NOT informed that...

- ...you have two tasks and make two reports (fruit and vegetable),
- ...how many points you can earn with any of the reports,
- ...how many points you can earn by reporting a specific item,
- ...whether you actually learned which item corresponds to which payoff before reporting.

Control questions

Please indicate whether the following statements are true or false:

Only one Observer will fill in a scorecard:

- ☐ True
☒ False

Observers learn that one item corresponds to a high payoff and seven items correspond to a low payoff:

- ☒ True
☐ False

The computer randomly determines whether Observers fill in my scorecard based on my report in the fruit or the vegetable task:

- ☒ True
☐ False

Observers know that I will report two items (fruit and vegetable):

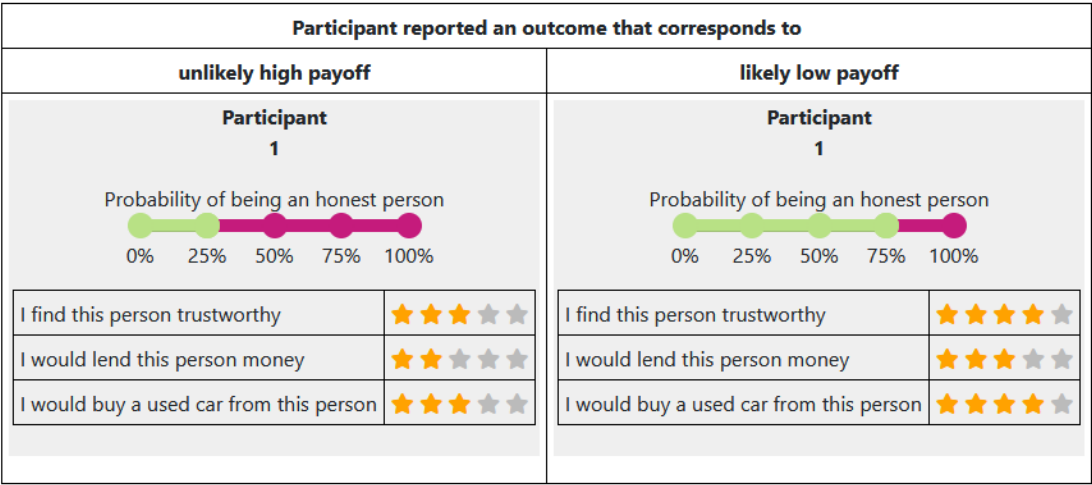
- ☐ True
☒ False

Next

Previous ratings

As explained, two Observers will reflect on your character and fill in your scorecard.

Since scorecards will not be shown to you, the figure below provides information on the most frequently chosen ratings by 27 Observers who reflected on participants' characters in a previous study and who considered participants' characters carefully (the graph excludes the three fastest Observers).



As becomes clear, Observers perceive participants whose reports resulted in the unlikely high payoff less favorably (e.g., less honest) than those whose reports resulted in more likely low paying outcomes.

Please indicate whether the following statement is true or false:

I can expect that the Observers will have a more favorable opinion of me if my report yields the unlikely high payoff:

- ☐ True
- ☒ False

Next

Ensuring correct representation of behavior

Recall that for one of the reports (vegetable or fruit) you may not know which item will result in which payoff. You may thus find yourself in the following scenario:

Potential scenario

It was randomly determined that two Observers fill in your scorecard based on your report in the vegetable task, and your vegetable report happens to coincide with the item that results in the unlikely high payoff of 1000 points, while at the same time your fruit report coincides with the low payoff.

Potential misperception of your character and favorable switch

In this scenario, the two Observers who fill in your scorecard based on your outcome in the (randomly selected) vegetable task may perceive your character unfavorably because the Observers only know that you obtained an unlikely high payoff but they neither know that there were two tasks nor whether you knew which payoff your report would result in.

For this reason, we offer you the option of a “favorable switch” if *all* of the following conditions apply:

- the vegetable task is the task based on which the Observers fill in your scorecard
- your report in the vegetable task results in the high payoff (1000 points)
- your report in the fruit task results in the low payoff (2 points).

That is, if you opt for the “favorable switch”, the Observers will rate your character based on the more likely, low payoff in the fruit task if the scenario described above arises.

Note: this switch will only be implemented if the exact scenario above arises (the randomly selected task for your scorecard is the vegetable task, and you obtained a high payoff from this report, and you obtained a low payoff from the fruit task).

Control questions

Please indicate whether the following statements are true or false:

I will now decide whether I want the favorable switch to be implemented in case the scenario described above arises:

- ☐ True
☐ False

The switch will only occur if my report in the fruit task coincides with one of the seven fruits that yield the low payoff:

- ☐ True
☐ False

Observers do not know that there are two tasks and thus they do not know how many points the vegetable and the fruit reports can yield:

- ☐ True
☐ False

Next

Your Choice

Recall: two Observers will judge your character.



Do you want to implement the favorable switch in case the scenario explained before arises?

☐ Yes ☐ No

Show scenario for favorable switch

Important: The two Observers will not be informed that you have to report both a fruit and a vegetable and will thus not learn whether you decided to switch which report (fruit or vegetable) your scorecard is based on.

Confirm

Payment for favorable switch

We have randomly determined a price that you need to pay for the favorable switch. If you are willing to pay that price, we will implement the switch for you. If you are not willing to pay the price, the switch will not be implemented.

The price we have randomly determined is one of the 17 possible prices shown below.

- Possible prices: 0, 25, 50, 75, 100, 125, 150, 175, 200, 250, 300, 350, 400, 500, 600, 700, 800

What is the highest price you are willing to pay for the favorable switch?

50 points ▼

The highest price you are willing to pay implies

- you are willing to pay any of the following prices: 0, 25, 50
- you consider all following prices too high: 75, 100, 125, 150, 175, 200, 250, 300, 350, 400, 500, 600, 700, 800

At the end of the study, we will check whether the price we have randomly determined is among the prices you are willing to pay. If so, the favorable switch (if applicable) will be implemented.

Note: You will only pay the price if the relevant scenario for the favorable switch arises.

Show scenario for favorable switch

The relevant scenario arises if all the following conditions apply:

- the vegetable task is the task based on which the Observers fill in your scorecard,
- your report in the vegetable task results in the high payoff (1000 points),
- your report in the fruit task results in the low payoff (2 points).

☒ Please confirm you are willing pay any of the following prices: 0, 25, 50

Next

Learning the payoff

You will now learn which fruit yields the higher payoff: **Grape**.

That is, if you report Grape you will receive 10 points. If you report any of the other fruits, you will receive 2 points.

Opportunity to learn the higher-paying vegetable before entering the report

By default, you will learn which vegetable yields the higher payoff right *after* entering your report.

You have the opportunity to pay a price to learn which vegetable yields 1000 points before entering your report.

If you decide to pay the price, we will subtract it from your total payoff and you will learn which vegetable yields 1000 points before entering your report. The next screen provides further information on the price.

Note: Observers do not know that you can choose to learn which vegetable yields 1000 points before reporting. Independent of your choice, Observers are only informed that you “may or may not” know the relationship of items to payoffs.

Control questions

Please indicate whether the following statements are true or false:

If I don't pay the price, I will learn the higher-paying vegetable right after entering my report:

- ☒ True
☐ False

Observers are informed about when I learn the relationship between my report and my payoff:

- ☐ True
☒ False

Next

Choice: revealing the higher-paying vegetable

We have randomly determined the price you need to pay to learn which vegetable yields 1000 points before entering your report. If you are willing to pay that price, you will learn which vegetable yields 1000 points before entering your report. If you are not willing to pay the price, you will learn which vegetable yields 1000 points right after entering your report.

The price we have randomly determined is one of the 16 possible prices shown below.

- Possible prices: 25, 50, 75, 100, 125, 150, 175, 200, 250, 300, 350, 400, 500, 600, 700, 800

What is the highest price you are willing to pay to learn which vegetable yields 1000 points before entering your report?

The highest price you are willing to pay implies

- you are willing to pay any of the following prices: 25, 50, 75, 100
- you consider all following prices too high: 125, 150, 175, 200, 250, 300, 350, 400, 500, 600, 700, 800

Before you report, we will check whether the price we have randomly determined is among the prices you are willing to pay. If so, you will learn which vegetable yields 1000 points before entering your report and pay the price we have randomly determined.

☒ Please confirm you are willing pay any of the following prices: 25, 50, 75, 100

Confirm

Information and Report

Task for scorecard and favorable switch

- The computer randomly determined that your scorecard will be based on your report in the following task: **vegetable**. Your scorecard will be based on the report in this task.
- The randomly determined price for the favorable switch was 25 points. As you indicated that you are willing to pay this price, we will implement the switch in case the relevant scenario arises, and your scorecard will then be based on your report in the fruit task.

Show scenario for favorable switch

The relevant scenario arises if all the following conditions apply:

- the vegetable task is the task based on which the Observers fill in your scorecard,
- your report in the vegetable task results in the high payoff (1000 points),
- your report in the fruit task results in the low payoff (2 points).

Decision to learn the higher-paying vegetable

The randomly determined price to learn which vegetable yields the higher payoff is 100 points.

You indicated that you are willing to pay this price to learn which vegetable yields the higher payoff before reporting. It is: **Cabbage**. That is, if you report Cabbage, you will receive 1000 points. If you report any of the other vegetables, you will receive 200 points.

Enter reports

Please now enter the randomly generated vegetable that resulted from the spin of the wheel of fortune:



☐ Tomato (200 points)



☐ Avocado (200 points)



☐ Carrot (200 points)



☐ Aubergine (200 points)



☐ Broccoli (200 points)



☐ Cabbage (1000 points)



☐ Mushroom (200 points)



☐ Pepper (200 points)

Recall: the higher-paying fruit is **Grape**.

That is, if you report Grape, you will receive 10 points. If you report any of the other fruits, you will receive 2 points.

Please now enter the randomly generated fruit that resulted from the spin of the wheel of fortune:



☐ Banana (2 points)



☐ Lemon (2 points)



☐ Apple (2 points)



☐ Grape (10 points)



☐ Pineapple (2 points)



☐ Watermelon (2 points)



☐ Strawberry (2 points)



☐ Cherry (2 points)

Confirm

Results

Your higher-paying vegetable is **Cabbage**.

Your higher-paying fruit is **Grape**.

- For your reported vegetable (Cabbage) you will receive 1000 points.
- For your reported fruit (Strawberry) you will receive 2 points.
- You paid 100 points to learn the higher-paying vegetable before reporting.
- The randomly determined price for the favorable switch was 25 points. As you indicated that you are willing to pay this price, we implemented the switch such that your scorecard will be based on the outcome in the fruit task. You will pay 25 points for the switch.

Your payoff is thus 877 points (= 1000 points + 2 points – 100 points – 25 points).

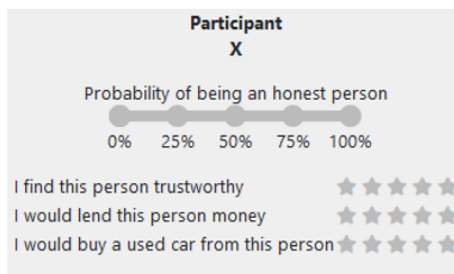
Next

Observers

Your task

On the following screens, you will receive information about the behavior of other Prolific participants who took part in our study. We call these participants "decision-makers".

Your task today is to rate the behavior of these decision-makers on a scorecard that includes four dimensions:



Your payoff

Your payoff depends on how well your ratings correspond with the ratings of another, randomly selected participant who also rates decision-makers. This participant has exactly the same task and has the same incentives that you have.

- There are four dimensions (probability that the decision-maker is an honest person, trustworthiness of the decision-maker, whether you would lend money to the decision-maker and whether you would buy a used car from the decision-maker)
- For each dimension, the scorecard has a 5-point-scale (honesty 0% 25% 50% 75% 100%, other ratings: 1-5 stars).
- You will receive 200 points for each of your ratings that exactly corresponds to the rating of the other participant you are paired with.
- As the scorecard has four dimensions, you may earn 0, 200, 400, 600, or 800 points in this task.

Control questions

Please indicate which option applies:

My task is to rate decision-makers on four dimensions:

- ☐ True
☐ False

I will earn 800 points for this task irrespective of my ratings:

- ☐ True
☐ False

If I have rated the decision-maker at the exact same level as the other participant I am paired with in three dimensions, but our ratings differ for the fourth dimension, I will earn:

- ☐ 800 points
☐ 600 points
☐ 400 points
☐ 200 points
☐ nothing

Next

Information about decision-makers

You will judge a report by decision-makers. For this report, decision-makers were asked to state the randomly generated item that a wheel of fortune had selected out of 8 possible items.

The reported item affected the decision-maker's payoff. One of the eight possible items resulted in a high payoff whereas the seven other items resulted in a low payoff.

Decision-makers may have learned the relationship between their report and the payoff before or after making their report.

Your task

- Reflect on what to conclude about the character of the decision-maker based on his or her report.
- Carefully design a scorecard, in which you rate the decision-maker on four dimensions as shown before.
- Openly express your opinion about what to conclude about a decision-maker depending on the report.

We will ask you to design the scorecard and express your opinion depending on whether the decision-maker's report resulted in an unlikely high or a likely low payoff.

Control questions

Please indicate which option applies:

Decision-makers are asked to report an item that they have chosen themselves:

- ☐ True
☐ False

In the decision-makers' task, an item has a chance of 1 in 8 to be realized if it is randomly generated by a wheel of fortune:

- ☐ True
☐ False

A randomly generated item is more likely to result in a high payoff than in a low payoff:

- ☐ True
☐ False

Next

How to rate

On the next page, we will ask you to rate the decision-maker. Please read the following instructions first.

Decision maker reported an outcome that corresponds to	
unlikely high payoff	likely low payoff
<p>Participant X</p> <p>Probability of being an honest person</p> <p>0% 25% 50% 75% 100%</p>	<p>Participant X</p> <p>Probability of being an honest person</p> <p>0% 25% 50% 75% 100%</p>
I find this person trustworthy	★★★★★
I would lend this person money	★★★★★
I would buy a used car from this person	★★★★★

Important: Consider both scenarios (report of an unlikely high-paying outcome vs report of a likely low-paying outcome) carefully. One of the scenarios will be randomly selected for each participant to determine his/her number of points. We will count the number of exact matches between your and the other participant's rating for the randomly selected scenario.

Feedback

At the end of this study, you will learn which of the two scenarios was randomly selected for your payments and see to what extent your ratings correspond to the ratings of the participant you are matched with.

We will also provide you with information on which fraction of the decision-makers falls into each scenario (based on their report).

We will also show you all ratings of your matched partner.

Control questions

Please complete the following statements:

I will be paid for my ratings in both scenarios:

- ☐ True
☐ False

I will be paid for one randomly selected scenario, depending on how well my ratings in this scenario correspond with those of the participant I am matched with:

- ☐ True
☐ False

Next

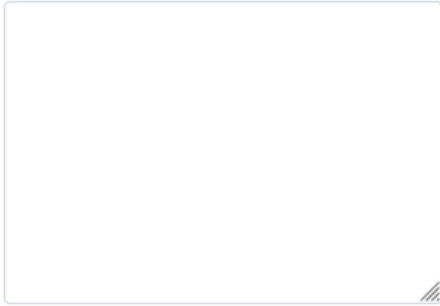
Please enter your rating below by selecting the values on both scorecards.

Decision maker reported an outcome that corresponds to	
unlikely high payoff	likely low payoff
<p>Participant X</p> <p>Probability of being an honest person</p> <p>0% 25% 50% 75% 100%</p>	<p>Participant X</p> <p>Probability of being an honest person</p> <p>0% 25% 50% 75% 100%</p>
I find this person trustworthy	★★★★★
I would lend this person money	★★★★★
I would buy a used car from this person	★★★★★

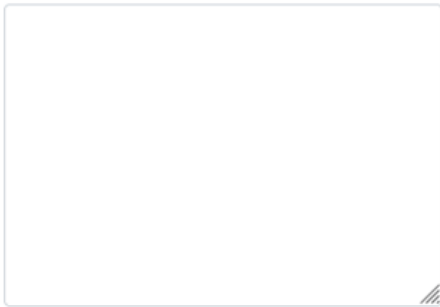
Next

Your opinion on behavior of decision-maker

1. You can openly express your opinion about the decision-makers who reported the **lower (more likely)** payoff here:



2. You can openly express your opinion about the decision-makers who reported the **higher (less likely)** payoff here:



Next

Your opinion on feelings of decision-maker

1. How comfortable do you think a decision-maker feels whose report results in the **unlikely, higher payoff**?

very uncomfortable very comfortable
☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐ 7

2. How comfortable do you think a decision-maker feels whose report results in the more **likely, lower payoff**?

very uncomfortable very comfortable
☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐ 7

Next

Questionnaire

Before the end of this study, we would like you to answer a few more questions.

Next

Questionnaire

Please indicate your gender:

Please indicate your age class:

.....

Approximately, how much money (in £) do you have available at the beginning of every month before paying expenses?

In politics, people often speak of "left" and "right" when it comes to denoting different political attitudes. Thinking about your own political views, where would you classify those views?

left right

☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐ 7

How would you rate your understanding of this study's instructions?

very low very high

☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐ 7

Is there anything else you would like to let us know?

Next

The ratings that will determine your payoff for this part were randomly chosen to be based on the outcome that results in an unlikely high payoff.

Participant X

Probability of being an honest person



50%

I find this person trustworthy

I would lend this person money

I would buy a used car from this person



Participant X

Probability of being an honest person



50%

I find this person trustworthy



I would lend this person money



I would buy a used car from this person



- Probability of being an honest person.
- I would buy a used car from this person.

- I find this person trustworthy.
- I would lend this person money.

Before proceeding to the final payment page, we will show you which decision-makers fall into which category.

78

Part 1 - Overview

Below you see which participants reported outcomes that yielded the unlikely high payoff, and which participants reported outcomes that resulted in the likely low payoff (the letter indicates the participation date and time, and the number the participant's identifier).

	unlikely high payoff	likely low payoff												
IDs	A-17, A-24, A-32, A-37, A-9, B-1, B-124, B-129, B-13, B-148, B-156, B-165, B-172, B-177, B-28, B-29, B-36, B-40, B-48, B-5, B-69, B-89, B-9, B-97, C-13, C-25, C-33, C-36, C-40, C-48, C-57, D-21, D-25, D-4, D-44, D-49, D-53, D-57, D-9, E-12, E-20, E-29, E-33, E-5, E-8, F-28, F-32, F-37, F-44, F-48, G-21, G-25, G-29, G-33, G-44, G-53, G-57, G-61, G-8, H-33, H-48, H-49, H-61, H-64, H-68, H-72, H-76, H-77, H-81, I-13, I-4, I-5, I-9, J-33, J-37, J-45, J-53, K-12, K-20, K-24, K-25, K-33, K-5, K-8	A-1, A-12, A-13, A-20, A-21, A-25, A-28, A-33, A-36, A-40, A-41, A-44, A-45, A-5, A-8, B-100, B-101, B-104, B-105, B-108, B-109, B-112, B-117, B-120, B-121, B-125, B-128, B-132, B-136, B-137, B-141, B-144, B-145, B-149, B-152, B-153, B-157, B-16, B-160, B-161, B-164, B-168, B-169, B-17, B-173, B-176, B-180, B-181, B-184, B-20, B-21, B-32, B-33, B-37, B-4, B-41, B-44, B-45, B-53, B-56, B-57, B-60, B-61, B-64, B-65, B-68, B-72, B-73, B-76, B-77, B-8, B-80, B-81, B-84, B-85, B-88, B-92, B-96, C-12, C-16, C-17, C-21, C-28, C-29, C-32, C-41, C-45, C-49, C-56, C-60, C-65, C-68, C-69, C-9, D-13, D-17, D-20, D-28, D-32, D-33, D-36, D-37, D-40, D-41, D-48, D-52, D-60, D-8, E-13, E-16, E-21, E-24, E-25, E-28, E-32, E-36, E-37, E-4, E-40, E-44, E-45, E-52, E-53, E-57, E-9, F-12, F-16, F-17, F-20, F-21, F-24, F-25, F-29, F-33, F-36, F-4, F-40, F-41, F-49, F-5, F-52, F-56, F-8, F-9, G-12, G-13, G-17, G-20, G-24, G-28, G-36, G-4, G-41, G-45, G-49, G-5, G-52, G-56, G-9, H-12, H-13, H-16, H-17, H-20, H-21, H-24, H-28, H-29, H-32, H-36, H-37, H-40, H-41, H-44, H-45, H-53, H-56, H-57, H-60, H-65, H-69, H-73, H-8, H-80, H-9, I-1, I-12, I-8, J-16, J-20, J-21, J-24, J-25, J-29, J-32, J-36, J-4, J-40, J-41, J-48, J-49, J-8, J-9, K-1, K-13, K-16, K-17, K-21, K-29, K-32, K-36, K-37, K-4, K-40, K-41, K-9												
Your ratings	<div><div>Participant see above for IDs</div><div>Probability of being an honest person</div><div><div></div><div></div></div><div>50%</div></div> <table><tr><td>I find this person trustworthy</td><td>★★★★★</td></tr><tr><td>I would lend this person money</td><td>★★★★★</td></tr><tr><td>I would buy a used car from this person</td><td>★★★★★</td></tr></table>	I find this person trustworthy	★★★★★	I would lend this person money	★★★★★	I would buy a used car from this person	★★★★★	<div><div>Participant see above for IDs</div><div>Probability of being an honest person</div><div><div></div><div></div></div><div>75%</div></div> <table><tr><td>I find this person trustworthy</td><td>★★★★★</td></tr><tr><td>I would lend this person money</td><td>★★★★★</td></tr><tr><td>I would buy a used car from this person</td><td>★★★★★</td></tr></table>	I find this person trustworthy	★★★★★	I would lend this person money	★★★★★	I would buy a used car from this person	★★★★★
I find this person trustworthy	★★★★★													
I would lend this person money	★★★★★													
I would buy a used car from this person	★★★★★													
I find this person trustworthy	★★★★★													
I would lend this person money	★★★★★													
I would buy a used car from this person	★★★★★													
The ratings of the participant you have been paired with	<div><div>Participant see above for IDs</div><div>Probability of being an honest person</div><div><div></div><div></div></div><div>50%</div></div> <table><tr><td>I find this person trustworthy</td><td>★★★★★</td></tr><tr><td>I would lend this person money</td><td>★★★★★</td></tr><tr><td>I would buy a used car from this person</td><td>★★★★★</td></tr></table>	I find this person trustworthy	★★★★★	I would lend this person money	★★★★★	I would buy a used car from this person	★★★★★	<div><div>Participant see above for IDs</div><div>Probability of being an honest person</div><div><div></div><div></div></div><div>50%</div></div> <table><tr><td>I find this person trustworthy</td><td>★★★★★</td></tr><tr><td>I would lend this person money</td><td>★★★★★</td></tr><tr><td>I would buy a used car from this person</td><td>★★★★★</td></tr></table>	I find this person trustworthy	★★★★★	I would lend this person money	★★★★★	I would buy a used car from this person	★★★★★
I find this person trustworthy	★★★★★													
I would lend this person money	★★★★★													
I would buy a used car from this person	★★★★★													
I find this person trustworthy	★★★★★													
I would lend this person money	★★★★★													
I would buy a used car from this person	★★★★★													
Participants' final scorecards based on your and your partner's rating	<div><div>Participant see above for IDs</div><div>Probability of being an honest person</div><div><div></div><div></div></div><div>50%</div></div> <table><tr><td>I find this person trustworthy</td><td>★★★★★</td></tr><tr><td>I would lend this person money</td><td>★★★★★</td></tr><tr><td>I would buy a used car from this person</td><td>★★★★★</td></tr></table>	I find this person trustworthy	★★★★★	I would lend this person money	★★★★★	I would buy a used car from this person	★★★★★	<div><div>Participant see above for IDs</div><div>Probability of being an honest person</div><div><div></div><div></div></div><div>75%</div></div> <table><tr><td>I find this person trustworthy</td><td>★★★★★</td></tr><tr><td>I would lend this person money</td><td>★★★★★</td></tr><tr><td>I would buy a used car from this person</td><td>★★★★★</td></tr></table>	I find this person trustworthy	★★★★★	I would lend this person money	★★★★★	I would buy a used car from this person	★★★★★
I find this person trustworthy	★★★★★													
I would lend this person money	★★★★★													
I would buy a used car from this person	★★★★★													
I find this person trustworthy	★★★★★													
I would lend this person money	★★★★★													
I would buy a used car from this person	★★★★★													
Percent of participants in this category	28%	72%												

Payoff information

Your participation payment is £1.50.

Your payoff from the scorecards is 400 points and thus equals £1.00.

You answered 8 out of 8 control questions correctly. For everyone who answered at least 5 control questions, we will pay the bonus. You will thus receive the bonus for answering control questions.

Your total payoff from this study is £2.70 (£1.50 + £1.00 + £0.20).

Please click on the following link to return to Prolific: <https://app.prolific.co/submissions/complete?cc=5CC3FEAC>

E.2 Validation tasks

E.2.1 Mind game

General Instructions

In this part, you will be shown a list of six African cities. Your task is to pick one of these cities and memorize it.

[NEXT]

List of African Cities

The following list shows six African cities. Please pick one of these cities and memorize it (the exact spelling is not important) [Order of names randomized on individual level, as on previous page; single choice]:

- Conakry
- Harare
- Luanda
- Malabo
- Praia
- Yamoussoukro

☐ I confirm that I have memorized one of the above cities

[NEXT]

Entry

[Association of cities and payoffs randomized on individual level.] Please now report the city you memorized. Depending on the city you memorized, you will earn between 0 points and 1250 points. The table below shows the payoff that corresponds to each possible report. I memorized the African city:

- Conakry
- Harare
- Luanda
- Malabo
- Praia
- Yamoussoukro

Name of City	Harare	Praia	Luanda	Malabo	Conakry	Yamoussoukro
Payoff	250 points	500 points	750 points	1000 points	1250 points	0 points

Table E.10: Example of mapping of reports to payoffs in mind game

[NEXT]

Payoff information

Thank you for your report. You reported [city] as your memorized city. By the way: [city] is the capital of [country]. Your payoff from this part is [payoff] points.

E.2.2 Sender-receiver game

General Instructions [Sender]

During this part, you will be matched with another participant whose choices do not affect your payoff. This participant needs to choose one out of ten available options (Option A to J), and the option the other participant chooses determines the other participant's payoff. However, the other participant does not know the payoff associated with each option. In contrast, you know which option will result in which payoff for the other participant and you can advise the other participant by sending a message of the form "Option X yields the highest possible payoff for you." While the other participant's payoff is determined by the option the other participant chooses, your payoff is determined by the message you send:

- If you send the message "Option [A] yields the highest possible payoff for you." to the other participant, you will earn 100 points.
- If you send any other message, you will earn 900 points.

You choose which message to send to the other participant. The computer randomly determines whether your message will be delivered. Then, the other participant chooses an option. The following table shows the ten possible messages you can send, the payoff each message yields for you, and the other participant's payoff if they follow your message.

Message you send the other participant	Your payoff if you send this message	Other participant's payoff if they follow your message
Option A yields the highest possible payoff for you.	100	900
Option B yields the highest possible payoff for you.	900	100
Option C yields the highest possible payoff for you.	900	100
Option D yields the highest possible payoff for you.	900	100
Option E yields the highest possible payoff for you.	900	100
Option F yields the highest possible payoff for you.	900	100
Option G yields the highest possible payoff for you.	900	100
Option H yields the highest possible payoff for you.	900	100
Option I yields the highest possible payoff for you.	900	100
Option J yields the highest possible payoff for you.	900	100

Two scenarios [Sender]

There are two possible scenarios, and in both scenarios, the other participant does not know which option yields which payoff for them. Further, the other participant neither learns your payoff, nor how it is determined. The two scenarios differ with respect to whether the other participant knows which payoff levels they can obtain.

Scenario 1

- The other participant does not know which possible payoff levels exist.
- The only information the other participant receives is the message.
- That is, the other participant cannot know whether the message is true or not.

Scenario 2

- The other participant is informed that two possible payoff levels exist, and that one option yields a payoff of 900 points while the nine other options yield a payoff of 100 points.
- In addition, the other participant receives the message.
- That is, the other participant may infer whether the message is true or not.

The computer has randomly determined whether Scenario 1 or Scenario 2 is payoff-relevant. Both scenarios were equally likely to be selected. However, you do not know which scenario is relevant. We thus ask you to select the message to be sent for both scenarios. After you have submitted your choice for each scenario, you will learn which scenario was relevant. Note: The other person does not know that there are two possible scenarios and receives the information based on the randomly determined payoff-relevant scenario.

[NEXT]

Choice [Sender]

The computer has randomly determined the relevant scenario. In the table below, please choose a message that will be sent in case the relevant scenario is:

Scenario 1:

- The other participant does not know which possible payoff levels exist.
- The only information the other participant receives is the message.
- That is, the other participant cannot know whether the message is true or not.

[TABLE WITH SELECTION OPTIONS AS SHOWN IN TABLE ABOVE]

Please select the message you would like to send

[CONFIRM]

Scenario 2:

- The other participant is informed that two possible payoff levels exist, and that one option yields a payoff of 900 points while the nine other options yield a payoff of 100 points.
- In addition, the other participant receives the message.
- That is, the other participant may infer whether the message is true or not.

[TABLE WITH SELECTION OPTIONS AS SHOWN IN TABLE ABOVE]

Please select the message you would like to send

[CONFIRM]

Payoff information [Sender]

The computer randomly determined scenario [SCENARIO] to be the relevant scenario. For this scenario, you chose the message "Option [CHOICE] yields the highest possible payoff for you". Your payoff for this part thus equals [Payoff] points.

[NEXT]

General instructions [Receiver]

During this part, you will be matched with another participant who can advise you on which out of ten available options to choose (Option A to J). The option you choose will determine your payoff. However, you do not know the payoff associated with each option when choosing an option. [if Scenario 2, add: One option yields 900 points while the other nine options yield 100 points.] The participant you are matched with knows to which payoff each option corresponds and can advise you by sending you one out of ten possible messages of the form: "Option X yields the highest possible payoff for you" where "X" will correspond to one out of the ten Options "A" to "J".

[NEXT]

Choice [Receiver]

The other participant sent you the following message:

Message [CHOICE SENDER]: "Option [CHOICE SENDER] yields the highest possible payoff for you."

Please select one option.
[CONFIRM]

Payoff information [Receiver]

You chose option [CHOICE RECEIVER]. This option is associated with a payoff of [Payoff] points for you. Your payoff for this part thus equals [Payoff] points. [NEXT]

E.2.3 Knowledge reporting task

Introduction

This part consists of two tasks.

Task Y:

On the following screen, you will see a list of 20 entries for 60 seconds. We will ask you how many of the entries you identify as important people in human history. You have to submit an answer within 60 seconds to receive 500 points for this task. If you do not submit an answer within that time, you will receive 0 points for this task.

After the task is completed, observers will be shown the number of people you identified, along with the number of entries all other participants identified. [NEXT]

Task Y

How many of the following entries do you identify as important people in human history?

- George Washington
- Leonardo da Vinci
- Nelson Mandela
- Marie Curie
- Cleopatra
- Mahatma Gandhi
- Makalalo Nihambo
- Augusta Cincoquanta
- Salazar Ariza
- Maria Heinrich von Graubert
- Chen-Shun Li
- Matteo Sacrosanto
- Natalia Krusajeva
- N'koto Maboyo
- Theodore Lomney II
- Henri Marques du Cressant
- Jonathan Hammersmith
- Takeo Nokosato
- Olivia Broker
- Mustafa Al-Jazaa

I identify the following number of entries as important people in human history:

Task Z

On the following screen, you will see a list of 20 entries for 60 seconds. We will ask you how many of the entries you identify as birds. You have to submit an answer within 60 seconds to receive 500 points for this task. If you do not submit an answer within that time, you will receive 0 points for this task.

After the task is completed, observers will be shown the number of birds you identified, along with the numbers all other participants identified. [NEXT]

Task Z

How many of the following entries do you identify as birds?

- Swan
- Duck
- Flamingo
- Owl
- Eagle
- Crow
- Plantipallo
- Coughting
- Nashvine
- Tuttlebeek
- Nimbar
- Cuttlesplint
- Nevalus Pantae
- Frogpiper
- Yellow Splintling
- Brown Belster
- Yukzi
- Membinta
- Cinpalla
- Katchuk

I identify the following number of entries as birds:

Payoff

- [if no humans timeout] You identified the important people in human history in time and will thus receive 500 points.
- [if humans timeout] You did not identify the important people in human history in time, so you will not receive points for this task.
- [if no birds timeout] You identified the birds in time and will thus receive 500 points.
- [if birds timeout] You did not identify the birds in time, so you will not receive points for this task.

Your total payoff for part 4 is [0/500/1000] points.

E.3 Survey questions

The following displays the vignette and questions for the unincentivized survey questions.

Please imagine the host of a live radio show calls you to participate in a radio raffle. Your task is the following: you must flip a coin 4 times and report the number of “tails” you flipped. You will receive 10 pounds for each reported “tail”. You know that there is no way the host can verify whether you reported the true number of “tails” you flipped.

Situation 1: *Imagine the unlikely case occurs that you flip 4 tails.*
How many tails will you report to the host? [Choice 0/1/2/3/4]

Situation 2: *Imagine the unlikely case occurs that you flip 0 tails.*
How many tails will you report to the host? [Choice 0/1/2/3/4]