

The Truth About Lies: A Meta-Analysis on Dishonest Behavior

Philipp Gerlach

Max Planck Institute for Human Development, Berlin, Germany

Kinneret Teodorescu

Technion - Israel Institute of Technology

Ralph Hertwig

Max Planck Institute for Human Development, Berlin, Germany

Over the past decade, a large and growing body of experimental research has analyzed dishonest behavior. Yet the findings as to when people engage in (dis)honest behavior are to some extent unclear and even contradictory. A systematic analysis of the factors associated with dishonest behavior thus seems desirable. This meta-analysis reviews four of the most widely used experimental paradigms: sender–receiver games, die-roll tasks, coin-flip tasks, and matrix tasks. We integrate data from 565 experiments (totaling $N = 44,050$ choices) to address many of the ongoing debates on who behaves dishonestly and under what circumstances. Our findings show that dishonest behavior depends on both situational factors, such as reward magnitude and externalities, and personal factors, such as the participant’s gender and age. Further, laboratory studies are associated with more dishonesty than field studies, and the use of deception in experiments is associated with less dishonesty. To some extent, the different experimental paradigms come to different conclusions. For example, a comparable percentage of people lie in die-roll and matrix tasks, but in die-roll tasks liars lie to a considerably greater degree. We also find substantial evidence for publication bias in almost all measures of dishonest behavior. Future research on dishonesty would benefit from more representative participant pools and from clarifying why the different experimental paradigms yield different conclusions.

Public Significance Statement

Reports on corruption in industry and politics, fake news, and alternative facts highlight how crucial honesty is to the functioning of societies. But what aspects make people act dishonestly? We review 565 experiments that tempted 44,050 participants to behave dishonestly. We show that the degree and the magnitude of dishonesty depend on properties of the person (e.g., age, gender) and the context (e.g., the incentive to misreport, the experimental setup).

Keywords: cheating, deception, dishonesty, ethical behavior, meta-analysis

Supplemental materials: <http://dx.doi.org/10.1037/bul0000174.supp>

The Enron accounting scandal, WorldCom’s Ponzi scheme, FIFA’s web of corruption, the Volkswagen Dieselgate emissions scandal, the Petrobras and Odebrecht bribery cases in Brazil—these are just a few examples of widespread dishonesty and fraud worldwide. According to Transparency International’s annual global survey of corruption levels, over two thirds of the 176

countries and territories in the Corruption Perceptions Index, 2016 fell below the midpoint of the scale from 0 (*highly corrupt*) to 100 (*very clean*; Transparency International, 2017). The global average score is a measly 43, indicative of endemic corruption. It seems that dishonesty is a widespread phenomenon. Against this background, it is not surprising that dishonesty has become a research topic in its own right. How can it be that so many seemingly normal and well-adjusted people behave dishonestly to such an extent that their behavior gravely harms others? Will anybody succumb to dishonesty in the “right” situation, or is there a dishonest personality type? These and related questions have a long tradition in experimental research, which has sought to reveal the dynamics behind the dark sides of human nature—such as obedience to authority or gawking bystanders’ failure to render help—by running highly controlled behavioral experiments (e.g., Darley & Latané, 1968; Milgram, 1974).

Within the last decade, experiments examining the prevalence and magnitude of dishonesty, as well as its enabling conditions, have generated a large body of empirical findings across several

Philipp Gerlach, Center for Adaptive Rationality, Max Planck Institute for Human Development, Berlin, Germany; Kinneret Teodorescu, Faculty of Industrial Engineering and Management, Technion - Israel Institute of Technology; Ralph Hertwig, Center for Adaptive Rationality, Max Planck Institute for Human Development.

We thank Susannah Goss and Anita Todd for editing our manuscript. We are grateful to the authors of the primary experiments for sharing their data and answering our queries.

Correspondence concerning this article should be addressed to Philipp Gerlach, Center for Adaptive Rationality, Max Planck Institute for Human Development, Lentzeallee 94, 14195 Berlin, Germany. E-mail: pgerlach@mpib-berlin.mpg.de

disciplines, including behavioral economics (Fischbacher & Föllmi-Heusi, 2013; Gneezy, 2005), neuroscience (Greene & Paxton, 2009), and psychology (Mazar, Amir, & Ariely, 2008). Unlike past research using qualitative case studies and/or surveys to measure self-reported dishonesty, this new line of research harnesses the tool of experimentation to quantify manifestations of dishonest behavior. Despite marked differences in their experimental details (Rosenbaum, Billinger, & Stieglitz, 2014), these experiments all generated a basic conflict between the temptation to behave dishonestly and the capacity to resist that temptation.

Several reviews of when and why people engage in (dis)honest behavior have already been published (Gino, 2015; Gino & Ariely, 2016; Jacobsen, Fosgaard, & Pascual-Ezama, 2018; Rasmussen, 2015; Rosenbaum et al., 2014). Although valuable and informative, these reviews have relied on narrative summaries, synthesizing the empirical findings on a case-by-case basis. The aim of this meta-analysis is to complement the narrative reviews by providing a systematic and statistical synthesis of the experimental findings. Meta-analyses have several advantages over narrative summaries, as will be outlined below. We therefore applaud recent endeavors to meta-analytically integrate empirical findings of dishonest behavior (Abeler, Nosenzo, & Raymond, 2016). In their quantitative review, Abeler, Nosenzo, and Raymond formalized a range of theories about when and why people engage in (dis)honest behavior, and they tested those theories against the data. Our meta-analysis extends their work by including additional experimental paradigms and by aiming to offer an analysis as comprehensive as possible, allowing us to pursue three broad goals.

Our first goal is to quantitatively synthesize the empirical findings on dishonest behavior yielded by four of the most popular experimental paradigms. Examining variations within and between the paradigms most commonly used to study dishonest behavior allows us to answer several open questions: Does the literature paint an adequate picture of the prevalence of dishonesty? Do studies with low power bias this picture? Do different experimental paradigms lead to different conclusions about the circumstances under which people behave dishonestly? Our quantitative answers are based on the combined empirical data of more than 100 experiments per experimental paradigm.

Our second goal is to examine personal and situational factors associated with dishonest behavior. Do greater rewards prompt more dishonest behavior? Do laboratory and online experiments yield similar conclusions about dishonest behavior? Are student samples representative for the degree of dishonesty in the greater population? Do men behave more dishonestly than women? Narrative reviews can provide useful summaries of the debates on situational and personal factors impacting dishonest behavior. Yet the quantitative nature of meta-analyses makes it possible to estimate the degree to which each factor promotes dishonest behavior. For example, by integrating even those articles that do not mention gender differences in our analyses, we provide a more unbiased evaluation of potential gender differences than narrative reviews can. Meta-analyses can even go beyond evaluating measured factors. For example, we examined the effect of the use of experimental deception on dishonest behavior among the participants. Experimental deception (experimenters deceiving their participants about aspects of the study) varied between studies but none of the primary investigations examined its moderating influence.

The third goal of our meta-analysis was to examine the interactions of the experimental paradigms with the personal and situational factors. The conclusions emerging from the literature as to when people engage in (dis)honest behavior are to some extent mixed, even contradictory. For example, whereas some experimental paradigms suggest that increasing the reward size leads to more dishonest behavior (Gneezy, 2005; Sutter, 2009), others suggest null effects (Fischbacher & Föllmi-Heusi, 2013; Gächter & Schulz, 2016; Hugh-Jones, 2016; Kajackaite & Gneezy, 2017; Mazar et al., 2008). Metaanalytical techniques make it possible to assess the circumstances under which the size of the temptation affects dishonest behavior. The results can thus inform future experimental research on the enabling and disabling conditions of dishonest behavior. Before presenting the findings, let us first introduce the four experimental paradigms included in this meta-analysis. We then summarize previous empirical findings, highlighting open questions and conflicting patterns of results. After outlining the data set, we introduce the standardized measures that allow us to classify and compare different types of dishonest behavior across and within the experimental paradigms.

How Is Dishonest Behavior Measured?

Experiments assessing dishonest behavior are relatively heterogeneous, thwarting easy comparison and replication (Rosenbaum et al., 2014). To be able to compile, compare, and analyze experiments, we focused on four of the most widely employed paradigms that assess dishonest behavior.

In *sender–receiver games* (Gneezy, 2005), two participants interact. One participant, the sender, learns about the payoffs of two or more options. The options differ in terms of how lucrative their payoffs are for the two participants. Yet the total earnings of the two players are typically a constant sum and one party's gain is the other's loss. The sender then decides which message she or he wants the receiver to read. For example, the sender can send a false message stating "Option A will earn you more money" or a true message stating "Option B will earn you more money." After reading the message, the receiver chooses between the options, not knowing the actual payoffs. However, it is the decision of the receiver that determines the one-off payoffs for both players. Senders thus face a dilemma between sending a truthful message and a false but potentially more lucrative message. The truthful message, if believed and acted upon by a receiver, will result in a lower payoff for the sender.

In *coin-flip tasks* (Buccioli & Piovesan, 2011), participants are asked to report private information on a randomly and self-generated dichotomous outcome, typically the result of a coin toss. Reporting one outcome (e.g., heads) wins participants some reward; reporting the other outcome (e.g., tails) leaves participants empty handed. The rate of dishonest reporting can be estimated only on the aggregate level. If data from a sufficiently large number of participants are combined, experimenters can compare the proportion of reported wins (e.g., 75% heads) to the theoretical baseline of randomly generated wins (e.g., 50% heads). In contrast to sender–receiver games, coin-flip tasks thus do not allow *individual* dishonest behavior to be directly observed or spotted. Moreover, participants typically interact with the experimenter rather than with another participant. What happens to the experimental money that is not paid out to participants is usually not specified:

the more people misreport, the more money is paid out (positive sum game). Hence, the negative consequences of dishonest behavior are less clear in coin-flip tasks than they are in sender–receiver games.

In *die-roll tasks* (Fischbacher & Föllmi-Heusi, 2013; Fischbacher & Heusi, 2008), participants are asked to report private information on a randomly generated continuous outcome, such as the roll of a die. Whereas the choice outcomes of sender–receiver games and coin-flip tasks are binary, die-roll tasks can have three or more possible outcomes per choice. Participants know in advance that each reported outcome is rewarded by a corresponding amount. For example, reporting a 1 pays \$1, a 2 pays \$2, a 3 pays \$3, and so forth. Like coin-flip tasks, die-roll tasks do not detect dishonest behavior at the individual level. Instead, experimenters can estimate the degree of dishonest behavior by comparing the mean outcome reported at the aggregate level (e.g., on average, people reported a score of 4.2) with a hypothetical, randomly generated distribution of outcomes (e.g., on average, a fair six-sided die would yield a score of 3.5). Similar to coin-flip tasks, die-roll tasks are typically positive sum games, in which the victim of dishonest behavior is not clearly defined and in which the more people cheat, the more money is paid out across the sample of participants.

In *matrix tasks* (Mazar et al., 2008), participants are typically presented with several matrices, each containing 12 three-digit numbers (e.g., 4.56). The task is to find the number pair in each matrix that adds up to exactly 10.00 (e.g., $4.56 + 5.44 = 10.00$). However, most participants do not manage to solve all matrices in the time allotted. When time has run out, participants are asked to count how many matrices they have solved and then to pay themselves accordingly (e.g., \$1 per solved matrix). Dishonest behavior in matrix tasks can be measured at either the aggregate level or the individual level. Matrix tasks that measure dishonest behavior at the aggregate level randomly assign participants to two groups: an experimental group and a control group. The answers of the experimental group are self-graded by the participants, whereas the answers of the control group are verified by the experimenter. The experimental group can thus cheat by inflating the number of allegedly solved matrices; the control group cannot. The amount of dishonest behavior can be estimated by comparing the total reported number of solved matrices from the experimental group with the actual number of solved matrices from the control group. Other matrix tasks allow dishonesty to be measured at the individual level. In this design, participants are given a collection slip in addition to the matrix sheet. After solving the matrices, they are asked to report the number of solved matrices on the collection slip. Unbeknownst to the participants, a unique identification code allows the matrix sheet to be matched to the collection slip. The experimenter thus has data on both the actual performance and the reported performance at the individual level. All matrix tasks, regardless of their design, use a continuous (rather than binary) outcome measure that allows the degree of dishonest behavior to be identified—similar to die-roll tasks. Matrix tasks, like coin-flip and die-roll tasks, are typically positive sum games, in which cheating has no identifiable victim. Unlike in the other experimental paradigms, in matrix tasks, the desire to appear honest seems also to be at odds with the desire to appear competent. Moreover, participants might falsely believe that they found the solution to a matrix although they did not, or they might miscount the total

number of “solved” matrices. False reporting in the matrix task should therefore not always be equated with dishonest behavior.¹

In sum, all four experimental paradigms involve information asymmetry and temptation: Participants know more than the people who determine their payoffs. False information, if believed, is rewarded. Table 1 provides an overview of the four experimental paradigms, highlighting some of their key differences.

Studies directly comparing two or more of the four experimental paradigms are rare. To the best of our knowledge, such comparisons are limited to comparing die-roll and matrix tasks (Gino, Krupka, & Weber, 2013; Gravert, 2013; Kajackaite, 2018). The comparisons suggest that people behave differently in die-roll and matrix tasks. However, the differences are not consistent: Whereas Gravert (2013) found that die-roll tasks yield more dishonesty behavior than matrix tasks, Kajackaite’s (2018) results suggest the opposite trend. Interestingly, die-roll and matrix tasks also regularly result in rather different estimates of the type of dishonest behavior: Whereas die-roll tasks typically find that at least some proportion of people improperly claim the maximum amount (maximal lying; e.g., Fischbacher & Föllmi-Heusi, 2013), it is often the case that few or none of the participants in matrix tasks lie to the full extent (truth stretching; Gino, Ayal, & Ariely, 2009; Mazar et al., 2008). Several theoretical reasons for these diverging findings are possible. For example, in the typical die-roll task, observing the most advantageous outcome is as probable as any other outcome (e.g., rolling a 6 is as likely as rolling a 1). In the matrix task, top performances and, by extension, top claims are potentially less likely than low or medium performances. Consequently, the risk of evoking disbelief by claiming high outcomes may seem smaller in die-roll tasks than in matrix tasks (see Gneezy, Kajackaite, & Sobel, 2018, for experimental evidence supporting this hypothesis). Alternatively, the social norms and psychological costs governing honesty may differ in domains of chance (die-roll tasks) versus skill (matrix tasks; cf. Cherry, Frykblom, & Shogren, 2002).

Die-roll and coin-flip tasks are often seen as two versions of the same experimental paradigm (e.g., Abeler et al., 2016; Jacobsen et al., 2018; Rosenbaum et al., 2014). However, the two paradigms differ as to whether participants can partially lie. That is, in die-roll tasks participants can somewhat inflate the observed outcome (truth stretching: e.g., reporting a 4 instead of the observed 3). In coin-flip tasks, participants who tossed the losing side and who decide to lie must go the full length by reporting the opposite of their observation. However, an increased distance between truth and temptation can cause people to refrain from lying (Hilbig & Hessler, 2013).

To the best of our knowledge, no study has directly compared the sender–receiver game with any of the other experimental paradigms. On the one hand, sender–receiver games may lead to more dishonest behavior because the opportunity to behave dishonestly is an explicit choice that is highly salient to the participant. In the other experimental paradigms potential liars must first become aware of their opportunity to misreport (Lohse, Simon, & Konrad, 2018). This would suggest that sender–receiver games are associated with more lying. On the other hand, sender–receiver

¹ We thank an anonymous reviewer for pointing out these causes of seemingly dishonest behavior in the matrix task.

Table 1
Typical Key Properties of the Four Experimental Paradigms

Property	Sender–Receiver	Coin flip	Die roll	Matrix
Dishonest behavior	Sending a false message to another participant	Misreporting a randomly generated outcome	Misreporting a randomly generated outcome	Misreporting one’s performance in a task
Measuring scale	Dichotomous	Dichotomous	Continuous	Continuous
Measuring level	Individual	Aggregate	Aggregate	Individual/aggregate
Total payoffs	Constant	Positive	Positive	Positive
Identifiable victim	Yes	No	No	No

Note. Total payoffs refer to the typical sum of all participants’ payoffs. In constant sum games, one participant’s gain is the other’s loss. Hence, the other participant is the identifiable victim of any act of dishonesty. In positive sum games, participants can earn more from being dishonest without inflicting a loss on another participant. Hence, there is no identifiable victim.

games may also lead to less lying for several reasons. First, sender–receiver games measure dishonesty at the individual level. Studies suggest that participants lie less if experimenters can observe the true outcome (Gneezy et al., 2018; Rosenbaum et al., 2014). Second, Sutter (2009) found that some honest senders do not expect receivers to follow their advice. Hence, some senders intend to mislead receivers by sending truthful messages, not believing that the receiver will follow the advice. The behavior is honest nonetheless. Third, the constant total payoffs common to sender–receiver games may cause senders to refrain from lying because inflicting losses on other participants—rather than the experimenter—may cause participants with social preferences to refrain from lying (e.g., people who care about the payoff of other participants; Gino et al., 2013; Gneezy, 2005). We return to discussing the potential effect of social preferences and payoffs below.

Overall, the four experimental paradigms may in fact measure different facets of dishonest behavior. Sending false messages to other participants (sender–receiver games), deviating from a randomly generated figure through inflating (die-roll) or through reversing the observed outcome (coin-flip tasks), or lying about one’s actual performance (matrix task) may represent psychologically distinct behaviors prompting different levels of dishonesty. The first goal of this meta-analysis is therefore to assess whether the experimental paradigms come to the same or different conclusions about dishonest behavior.

Acting Dishonestly: Theory and Data

The conventional economic model assumes that people are willing to misreport private information if the material incentives of acting dishonestly outweigh those of acting honestly (Becker, 1968). In theory, the prototypical *homo economicus* will engage in dishonesty—even to the maximal possible extent—whenever this behavior pays off. Yet experiments on dishonest behavior suggest that people often behave otherwise: First, people acting like *homo economicus* represent only a fraction of all observations. A substantial proportion of individuals behaves completely honestly despite material incentives (Abeler et al., 2016; Fischbacher & Föllmi-Heusi, 2013). Second, the majority of those who behave dishonestly do so only to the extent that they can appear honest (to themselves, in the form of internalized norms: Mazar et al., 2008; or to others, in the form of social norms: Abeler et al., 2016; Gneezy et al., 2018; relatedly, see: Batson, Kobrynowicz, Dinnerstein, Kampf, & Wilson, 1997; Dana, Weber, & Kuang, 2007;

Schweitzer & Hsee, 2002; Tenbrunsel & Messick, 2004). That is, individuals often shy away from claiming the maximum potential payoff. Third, the degree to which individuals engage in dishonest behavior largely depends on personal and situational factors.

Personal Factors

Empirical investigations of dishonest behavior have looked at populations ranging from Franciscan nuns (Utikal & Fischbacher, 2013) to maximum security prisoners (Cohn, Maréchal, & Noll, 2015). Some of the personal factors most frequently assessed include gender, age, student status, and study major. Also the recruitment platform (e.g., Amazon’s Mechanical Turk) can draw on different populations, which, in turn, can influence dishonest behavior although such sampling effects have rarely been examined (see Peer, Brandimarte, Samat, & Acquisti, 2017 for an exception).

Gender. There has been substantial debate on gender differences in dishonest behavior (Capraro, 2018). Initially, Dreber and Johannesson (2008) found that 55% of men but only 38% of women sent false messages in sender–receiver games. Yet, two replication attempts failed to find similar results (Childs, 2012b; Gylfason, Arnardottir, & Kristinsson, 2013). On a broader scale, empirical evidence on gender effects is not clear cut. Although most studies have concluded that men behave more dishonestly than women (e.g., Cappelen, Sørensen, & Tungodden, 2013; Conrads, Irlenbusch, Rilke, & Walkowitz, 2013; Friesen & Gangadharan, 2012; Holm & Kawagoe, 2010; Houser, Vetter, & Winter, 2012; Ruffle & Tobol, 2014), others have found no gender differences (e.g., Abeler, Becker, & Falk, 2014; Aoki, Akai, & Ono-shiro, 2013; Arbel, Bar-El, Siniver, & Tobol, 2014; Erat & Gneezy, 2012; Holm & Kawagoe, 2010; Lundquist, Ellingsen, Gribbe, & Johannesson, 2009). One study even indicated that women behave more dishonestly than men (Ruffle & Tobol, 2014). A recent meta-analysis on gender differences in sender–receiver games concluded that gender differences depend on the relation of reward size to externalities (Capraro, 2018).

Experimental studies on dishonesty rarely refer to theoretical reasons for gender differences. Studies in related fields, such as those on academic integrity and risk taking, have suggested that men fear the possible sanctions for being caught cheating less than women (Hendershott, Drinan, & Cross, 1999). Hence a higher propensity for social risk taking among men (Croson & Gneezy, 2009) or a different perception of the risk may explain why men behave more dishonestly than women, even in experiments that

never expose individuals as liars. Other gender differences are conceivable too. For example, women may have different other-regarding preferences, and therefore may refrain from dishonest behavior more strongly if other people have to bear the costs of their actions (Croson & Gneezy, 2009).

Age. With respect to age effects, some studies have found that younger participants behave more dishonestly than older participants (Conrads et al., 2013; Friesen & Gangadharan, 2013; Glätzle-Rützler & Lergetporer, 2015), yet others have failed to find age effects (Abeler et al., 2014; Bucciol & Piovesan, 2011; Conrads & Lotz, 2015; Gino & Margolis, 2011), and at least one study has indicated that older people are more dishonest (e.g., Friesen & Gangadharan, 2012). The reasons for age effects are unclear and little discussed in the literature. Theories on risk taking may be informative for the age effect too: Studies have suggested that the propensity for risk taking declines with age (e.g., Josef et al., 2016; Mata, Josef, & Hertwig, 2016). Alternative explanations are differences in social preferences. Older people could be more agreeable and conscientious than younger people (Roberts, Walton, & Viechtbauer, 2006).

Students versus nonstudents. Most experiments have relied on student samples; relatively few experiments have assessed the degree to which results obtained from student samples generalize to other population groups. Most studies have suggested that students behave more dishonestly than a more representative sample of the population (Abeler et al., 2014; Aoki et al., 2013; Fosgaard, Bucciol, Hansen, & Piovesan, 2018); only one study found no systematic differences between students and more representative participant groups (Gunia, Barnes, & Sah, 2014). The effect of student population may be confounded by other effects. For example, students are likely younger than the general population and, consequently, age effects may contribute to students behaving more dishonestly. Relatedly, correlates of being a student, such as cognitive abilities, are recognized as an important factor for dishonest behavior (Gino & Ariely, 2012; Ruffle & Tobol, 2017).

Economics/business major. Numerous studies have compared students majoring in economics and business with other students. Some found that economics and business majors behaved more dishonestly than other students (Childs, 2012a, 2013; Lewis et al., 2012; Lundquist et al., 2009); others reported interaction effects with experimental factors or null effects (Gino et al., 2013; Muñoz-Izquierdo, de Liaño, Rin-Sánchez, & Pascual-Ezama, 2014). A greater propensity for dishonesty among economics and business students might be explained through social norms theories, which argue that people are willing to bear the costs of socially desirable behavior (e.g., acting honest) if they believe that sufficiently many other people would do the same (Bicchieri, 2006; Kallgren, Reno, & Cialdini, 2000). Studies have shown that economics and business students are more skeptical about other's honesty (López-Pérez & Spiegelman, 2012). As a consequence, economics and business students may not feel obliged to behave normatively correctly (Gerlach, 2017). We return to more direct tests of social norms theories in the following section. Alternative explanations are that people who behave more selfishly are drawn to study economics or business or that economics students learn to maximize profits when possible (Frey, Pommerehne, & Gygi, 1993; Kirchgässner, 2005; Scott & Rothman, 1975).

Mechanical Turk. In recent years, experiments on dishonesty have increasingly relied on samples recruited through Amazon's crowdsourcing platform, Mechanical Turk (MTurk; e.g., Biziouvan-Pol, Haenen, Novaro, Liberman, & Capraro, 2015; Charness, Blanco, Ezquerro, & Rodriguez-Lara, 2018; Gino & Wiltermuth, 2014; Gunia et al., 2014; Hildreth, Gino, & Bazerman, 2016; Kouchaki & Smith, 2014; Lee, Im, Parmar, & Gino, 2015; Peer, Acquisti, & Shalvi, 2014; Wang & Murnighan, 2017). To the best of our knowledge, only one study compared the levels of dishonesty among MTurk workers to those from other online platforms (Peer et al., 2017). The findings suggest that MTurk workers behave more dishonestly than participants from other platforms. A potential reason is that a large proportion of MTurk workers are "professional survey takers" who complete experimental tasks on a daily basis and who are therefore less naïve than conventional participant samples (Peer et al., 2017). For example, some MTurk workers may know that they can misreport in die-roll tasks whereas more naïve participants may not immediately be aware of this opportunity.

Situational Factors

Situational factors examined to date include the influence of normative cues (e.g., honesty reminders), the investigative setting, externalities, and the magnitude of the potential reward on dishonest behavior. Experimental deception may also influence dishonest behavior among the participants, although its effect has yet to be examined.

Normative cues. A number of studies have sought to remind participants about ethical concepts (e.g., by making them recall the Ten Commandments; Mazar et al., 2008) or by exposing them to information about the (dis)honest behavior of other people (social information; e.g., by introducing a confederate who allegedly lied; Gino et al., 2009).² The rationale behind priming ethical concepts is that they make ethical norms salient, whereas social information makes breaking ethical norms acceptable. The results seem unambiguous and consistent across the experimental paradigms: Ethical reminders decrease dishonesty (Bucciol & Piovesan, 2011; Mazar et al., 2008; Shu, Gino, & Bazerman, 2011), whereas providing social information about the dishonesty of other people increases dishonesty (Fosgaard, Hansen, & Piovesan, 2013; Gunia, Wang, Huang, Wang, & Murnighan, 2012; Innes & Mitra, 2013; Welsh, Ellis, Christian, & Mai, 2014).

Investigative setting. Physical distance to the person one is lying to may increase dishonest behavior. In particular, it has been argued that impression management becomes more important the closer the social proximity (Ackert, Church, Kuang, & Qi, 2011). In die roll tasks, for example, extreme outcomes (e.g., 6s) are more frequently reported in online studies than in laboratory studies (Abeler et al., 2014; Conrads & Lotz, 2015). Previous examinations of investigative setting, however, have been limited to die-roll tasks.

Externalities. Dishonest behavior may be sensitive to the degree to which other people are harmed by or benefit from it (Erat & Gneezy, 2012; Faravelli, Friesen, & Gangadharan, 2015; Gino,

² Studies have suggested that individuals' beliefs about what other people do and what they think is socially desirable are interconnected (Eriksson, Strimling, & Coultas, 2015).

Ayal, & Ariely, 2013; Gneezy, 2005; Hurkens & Kartik, 2009; Muñoz-Izquierdo et al., 2014; Rigdon & D’Esterre, 2014; Wiltermuth, 2011). In sender–receiver games, for example, one participant’s gain is typically the other’s loss (constant sum and identifiable victim). In coin-flip tasks, die-roll tasks, and matrix tasks, the externalities are commonly less clear (positive sum and no identifiable victim). Following Gneezy (2005), a number of researchers have manipulated the extent to which the other participant is harmed by trusting a false message in sender–receiver games, thereby changing the typical constant sum aspect of the game (Erat & Gneezy, 2012; Hurkens & Kartik, 2009; Sutter, 2009; Wang & Murnighan, 2017). Relatedly, payoffs in some coin-flip tasks (e.g., Kajackaite & Gneezy, 2017; Muñoz-Izquierdo et al., 2014), die-roll tasks (e.g., Gino et al., 2013), and matrix tasks (e.g., Faravelli et al., 2015; Rigdon & D’Esterre, 2014) have been modified to constant sum games by imposing externalities on other participants (identifiable victim). Whereas findings on sender–receiver games suggest that greater externalities result in less dishonest behavior (Gneezy, 2005; Hurkens & Kartik, 2009), the behavioral consequences of externalities in other experimental paradigms are less clear. For example, introducing externalities in die-roll tasks (Abeler et al., 2016; Fischbacher & Föllmi-Heusi, 2013) or in coin-flip tasks (Kajackaite & Gneezy, 2017) did not seem to affect (dis)honest behavior.

Reward size. The reward for acting dishonestly is a central element of experiments on dishonesty. Somewhat counterintuitively, it has been argued that greater rewards might lead to less dishonest behavior because the psychological costs of cheating increase (Mazar et al., 2008). An alternative explanation may be that the fear of being exposed as a liar increases with stake size (Kajackaite & Gneezy, 2017). However, most empirical findings suggest that dishonesty is relatively independent of reward size (Abeler et al., 2016; Fischbacher & Föllmi-Heusi, 2013; Hugh-Jones, 2016; Kajackaite & Gneezy, 2017) or that greater incentives increase dishonesty (Conrads, Irlenbusch, Rilke, Schielke, & Walkowitz, 2014; Gneezy, 2005; Kajackaite & Gneezy, 2017). The rationale is that greater rewards represent greater temptations. Reward size may even have different, partly opposing effects for different individuals (Hilbig & Thielmann, 2017). For example, whereas some potential liars may fear that greater rewards lead to more shaming, others may find greater rewards more tempting.

Experimental deception. Participants are not the only ones who can behave dishonestly. Experimenters too can provide false information about key aspects of the experiment. For example, experiments may intentionally misinform participants about the purpose of the experiment (e.g., Cai, Huang, Wu, & Kou, 2015; Gino & Margolis, 2011; Gino & Mogilner, 2014; Gino, Norton, & Ariely, 2010; Gino & Wiltermuth, 2014; Gu, Zhong, & Page-Gould, 2013; Shu et al., 2011). Such forms of experimental deception³ are relatively widespread in behavioral research, especially in psychology (Hertwig & Ortmann, 2008a) relative to experimental economics (Hertwig & Ortmann, 2001). In general, the practice of experimental deception is often defended in terms of experimental control: Experimental deception makes it possible to devise environments where socially undesirable and possibly rare behaviors can emerge that would be less likely to arise naturally if people suspected the true purpose of the study (Bonetti, 1998; Kimmel, 2007). If so, experimental deception would make easier to observe dishonest behavior by creating an experimental

environment in which participants feel ‘safe’ to behave dishonestly. If noticed, however, experimental deception may actually harm experimental control, by inducing suspicion among the participants.

What are the possible effects of suspicion of experimental deception? It may lead to less dishonesty among the participants, for example, because participants (sometimes correctly) infer that their responses do not remain anonymous. People might therefore respond strategically and, being sensitive to reputation concerns, “bend over backwards” (Kimmel, 1996, p. 68) to present themselves as honest individuals (cf. Gneezy et al., 2018). Alternatively, one may on theoretical ground expect that suspicion of experimental deception triggers more, rather than less, dishonesty among the participants because suspicion induces a breakdown of ethical norms. For example, according to the broken-windows theory, signs of disorderly and petty criminal behavior trigger more such behavior, causing it to spread in the social environment (Keizer, Lindenberg, & Steg, 2008). More broadly, social norm theory (e.g., Bicchieri, 2006; Cialdini, Reno, & Kallgren, 1990) suggests that by interacting with others, people learn about the social norms in the environment and react accordingly. Thus, suspicion might undermine honesty norms (in the sense that if the experimenters use deception, then it is acceptable to follow suit). In addition, social exchange theory (e.g., Cropanzano & Mitchell, 2005) highlights the role of reciprocity in social situations in which actions are contingent on the behavior of other people (e.g., honest deeds are rewarded and dishonest deeds are punished). Therefore, participants might deceive more when they suspect experimental deception (e.g., in the sense that if experimenters deceive, participants will deceive them back or punish them by claiming more money from them). Experimental deception may also induce anger (Motro, Ordóñez, Pittarello, & Welsh, 2016), resulting in impulsive behavior (Anderson & Bushman, 2002; Denson, DeWall, & Finkel, 2012) and risk taking (Lerner & Keltner, 2001).

Overall, the empirical findings on personal and situational factors impacting dishonest behavior are partly mixed, to some extent contradictory, and, in the case of experimental deception, even unknown. A systematic analysis of the factors associated with dishonest behavior thus seems desirable. Moreover, there is a possibility of publication bias, that is, the selective reporting of experiments with significant effects only. Most experiments have relatively small sample sizes, raising the question of how robust the identified effects are. For example, gender differences in dishonesty may be more likely reported if the effect reaches conventional levels of significance. We address these and related issues systematically and quantitatively using metaanalytical techniques.

³ There is a substantial debate on what counts as deception and what does not (Hertwig & Ortmann, 2008a, 2008b). Here we define experimental deception in terms of what seems to be the common denominator among researchers: Deception is the intentional provision of false information; withholding potentially relevant information is not.

Method

Search

In January 2018, we searched the Google Scholar databases (scholar.google.com) for all scientific articles that cited the seminal investigations introducing the four experimental paradigms: Gneezy (2005) for sender–receiver games, Bucciol and Piovesan (2011) for coin-flip tasks, Fischbacher and Heusi (2008; or the later publication Fischbacher & Föllmi-Heusi, 2013) for die-roll tasks, and Mazar et al. (2008) for matrix tasks. Simultaneously, we sent calls inviting information about further published and unpublished literature to the mailing lists of the *Academy of Management* (Organizational Behavior), the *Economic Science Association*, the *European Association for Decision Making*, the *Society for Judgment and Decision Making*, and the *Society for Personality and Social Psychology*. The search and the call covered all journal articles, book chapters, working papers, discussion papers, and scientific theses in order to minimize potential bias arising from the publication of only significant results. In the following, we refer to each item as an article. Only one-shot, fully anonymous, and incentivized experiments, in which dishonest behavior could not be sanctioned, were included in our analysis. A detailed description of the inclusion criteria is given in Appendix A. Figure 1 provides an overview of the selection process and the number of articles identified per step.

For all articles that fulfilled the inclusion criteria, we contacted the authors to obtain the primary data. If authors were unable or unwilling to provide us with the primary data, we inferred the necessary information from the test statistics and/or figures provided. Altogether, our inclusion criteria resulted in a pool of 130 articles (with a total of $N = 44,050$ observations) covering 558 experiments: 165 sender–receiver games ($n = 9,243$), 163 coin-flip tasks ($n = 13,879$), 129 die-roll tasks ($n = 14,835$), and 101 matrix tasks ($n = 6,093$). A complete list of the studies included can be found in Appendix A.

Coding

Standardized report. One way of comparing dishonest behavior within and between the experimental paradigms is the standardized report (Abeler et al., 2016), M_r :

$$M_r = \frac{m - t}{t - t_{\min}} \text{ if } m < t \quad (1)$$

and

$$M_r = \frac{m - t}{t_{\max} - t} \text{ if } m \geq t \quad (2)$$

where m is the actual mean report per study, t is the expected report if participants were honest, t_{\min} is the minimum possible report, and t_{\max} is the maximum possible report. For example, in a sender–receiver game, t , t_{\min} and t_{\max} are 0%, 0%, and 100%, respectively. In a coin flip task with a 50:50 chance of winning, t , t_{\min} and t_{\max} would be 50%, 0%, and 100%, respectively. In a die-roll task with a six-sided die where each score point translates into \$1 t , t_{\min} , and t_{\max} would be 3.5, 1, and 6, respectively. In a matrix task, in which the control group (aggregate level) or the

participants (individual level) solved, on average, 5 out of 20 matrices t , t_{\min} , and t_{\max} would be 5, 0, and 20, respectively.

The standardized report M_r can range from 100%, indicating that everybody cheated to the maximal degree, to $-100%$, indicating that all participants (oddly) claimed the lowest possible reward. The standardized report includes 0%, indicating that participants reported honestly. For example, in a die-roll task with a six-sided die where each score point translates into \$1 and in which all participants reported rolling $m = 6$ would translate into $M_r = 100%$, indicating that everybody who could in theory cheat maximally cheated in practice; an average report of $m = 3.5$ would equal the expected report of a fair die roll and would thus translate into $M_r = 0%$, indicating that participants reported honestly ($m = t$); if all participants reported rolling the lowest possible score, $m = 1$, the standardized report would be $M_r = -100%$. Negative values of the standardized report indicate that participants (oddly) claimed less than would be expected from honest reporting of a representative outcome distribution. For example, if the winning side in a coin-flip task with a single coin was reported in “only” $m = 44%$ of tosses (expected: $t = 50%$), then the standardized report would be $M_r = -12%$ ($= [44\% - 50\%]/[50\% - 0\%] = [m - t]/[t - t_{\min}]$).⁴

The standardized report thus quantifies the percentage of people who behaved dishonestly (the rate of liars) and the level of their dishonest behavior in a single measure. We focused on this measure in our analyses, as it allowed us to conduct comparisons within and across the four paradigms without limiting the data set (see Appendix B for details). Note, however, that sender–receiver games and coin-flip tasks do not allow for differentiation in the degree of dishonest behavior. Here, the decision to behave dishonestly is an all-or-nothing one (dichotomous measurement scale).

⁴ It can be argued that negative standardized reports do not reflect dishonest behavior because there is no apparent reason to claim less than one truthfully observed. To avoid distortion in the distribution of the standardized report, we allowed for negative standardized reports rather than excluding these values or making them zero. This is because negative and positive standardized reports could be, to some degree, the result of random sampling errors. For example, in die-roll tasks, participants may observe low numbers by chance. Consequently, the standardized report takes a negative value although participants were honest. In the opposite case—that is, when participants observed high numbers by chance—the standardized report may take a positive value although participants were honest. If the standardized report is allowed to take both positive and negative values, such random fluctuations cancel each other out with sufficient observations. The standardized report in most coin-flip and most die-roll tasks is symmetric, such that $t = (t_{\max} - t_{\min})/2$ and thus $|t - t_{\min}| = |t_{\max} - t|$ and $M_r = (m - t)/[(t_{\max} - t_{\min})/2]$. This does not hold for most matrix-tasks, however. For example, in a matrix task in which the control group solved $t = 5$ of a maximum of $t_{\max} = 20$ matrices (and $t_{\min} = 0$) an actual claim of $m = 2$ matrices would lead to $M_r = (2 - 5)/(5 - 0) = -60%$. In contrast, a mean claim of $m = 8$ matrices would lead to $M_r = (8 - 5)/(20 - 5) = 20%$. Thus despite that claiming 2 and 8 matrices are symmetric about 5 the claims result in asymmetric M_r . When calculating the grand mean of M_r , such asymmetry could introduce a bias (across all matrix tasks, the grand mean was $t = 7$ and thus less than $t_{\max}/2$ because the grand mean of $t_{\max} = 20$). Note, however, that only a single study had a negative M_r (Mazar et al., 2008, study 1; see Table A4 for an overview on the values of M_r). This study was excluded from most analyses because it included an honesty reminder. It is nonetheless possible that the study may have biased the regression analyses in Table 4. Appendix C therefore tests the robustness of the regression analysis of Table 4 by replacing the outcome measure. The results were qualitatively similar. We thank an anonymous reviewer for drawing our attention to the issue of asymmetry.

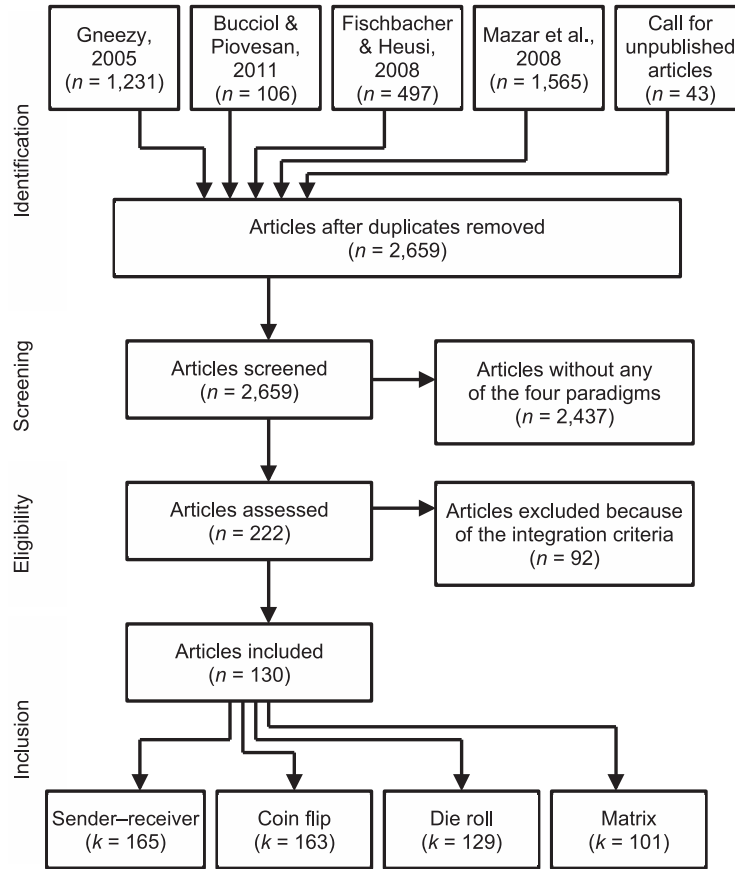


Figure 1. PRISMA (preferred reporting items for systematic reviews and meta-analyses) flow diagram describing the article search and selection process. All articles produced or published by the end of January 2018, with n = number of identified articles; k = number of identified experiments. Each experimental condition was counted as one experiment. In each experiment, actual responses were compared with a hypothetical distribution of honest answers (e.g., a hypothetical distribution of fair die rolls) or, in the case of aggregate-level matrix tasks, with the actual performance level of a control group in which dishonesty was impossible.

Hence, for sender–receiver games and most⁵ coin-flip tasks, the standardized report directly corresponds to the percentage of tempted people (e.g., participants who presumably observed the losing side) acting dishonestly. For example, a standardized report of $M_r = 23\%$ in sender–receiver games means that 23% of senders chose to convey the false message. For coin flip games, $M_r = 23\%$ means that 23% of the participants who presumably observed the losing side reported the winning side. Moreover, the standardized report in sender–receiver games cannot possibly take negative values. Die-roll and matrix tasks, in contrast, have a continuous scale. Here, the standardized reports do not distinguish between the percentage of people who behaved dishonestly (the rate of liars) and the degree of their dishonest behavior. For example, in a matrix task a standardized report with $M_r = 30\%$ indicates that 30% of unsolved matrices were claimed as solved. This could result from a few people misreporting to a high degree (low rate of liars with high level of dishonesty) or from many people misreporting to a lower degree (high rate of liars with low level of dishonesty; see [Appendix C](#) for more detailed analyses in tasks with continuous outcome measures). To disentangle the percentage of dishonest people from the degree of dishonesty, we examined an additional measure of dishonest behavior.

Rates of liars. The rate of liars, M_{liars} , indicates the percentage of participants who acted dishonestly, irrespective of the degree of their dishonest behavior. In the two paradigms with dichotomous scales (sender–receiver and coin-flip tasks), the rate of liars is practically equivalent to the standardized report. In the two paradigms with continuous scales (die-roll and matrix tasks), the rate of liars indicates the proportion of participants who claimed more than they were eligible for—regardless of how much more they claimed. To estimate the rate of liars in coin-flip and die-roll tasks, we used only a fraction of all observations. For die-roll tasks, the rate of liars was calculated using only reports of the lowest possible outcome. We assumed that the lowest possible outcome was reported only by participants who truly observed it. Whereas reporting all other outcomes may be the result of a self-serving lie (e.g., observing a 1 but reporting a 3), reporting the lowest possible outcome can be assumed to occur only when participants truly observe it. For example, a six-sided die is ex-

⁵ Some coin-flip tasks used randomly generated outcomes with a chance of winning other than 50:50. The standardized report and the rate of liars can therefore actually diverge for some coin-flip tasks (see [Appendix B](#)).

pected to yield the lowest pip in about 17% of the rolls. Hence, if “only” 8% of the participants reported the lowest pip the rate of liars converts to 53% (see Fischbacher & Föllmi-Heusi, 2013, for a proof). Because participants who observed the lowest outcome were maximally tempted to lie, the estimated rate of liars is expected to be an upper bound to the “true” rate of liars in die-roll tasks (see Appendix B for further details). For matrix tasks, the rate of liars was calculated only for experiments that measured individual-level behavior. For calculating the rate of liars in die-roll and matrix tasks, we needed access to the primary data. These methodological constraints reduced the number of observations and the number of experiments eligible for calculating the rate of liars (see Appendix B for detailed calculations). In the main analyses, we therefore focus on the standardized report, which uses all observations.

To combine the measures of dishonest behavior, we used multilevel models, which permit for random variations within studies (level 1; e.g., treatment 1 vs. treatment 2 of the same study) and between studies (level 2; e.g., Study 1 and Study 2). To quantify the variations in the effect size distributions we provide the I^2 statistic, which is the between-study variance independent of the number of experiments, and the τ^2 statistic, which is the estimated variance of underlying effects across studies (Huedo-Medina, Sánchez-Meca, Marín-Martínez, & Botella, 2006). If studies with greater standard errors yield more extreme outcome measures, publication bias seems likely. To identify and counteract the risk of selective reporting of only significant results (publication bias), we used various quantitative methods that assume different forms and severity of publication bias. To detect publication bias, Egger, Davey Smith, Schneider, and Minder (1997) suggested regressing the outcome measure (i.e., M_r , M_{liars}) against the standard error. A significant beta weight then serves as an indicator of publication bias, which can be addressed through several “correction” methods with different assumptions. We used the relatively common methods of trim and fill (Duval & Tweedie, 2000) and Vevea and Hedges’s (1995) weighting function. Trim and fill adds hypothetical experiments to the analysis until the observations are symmetrically distributed around the average effect size. Vevea and Hedges’s weighting function adjusts effect sizes based on the assumption that results were biased because of one-tailed p values. It is worth noting that the different “correction” methods can lead to dissimilar estimates of the adjusted effect size (McShane, Böckenholt, & Hansen, 2016).

For all analyses, we used the statistical software *R* (R Development Core Team, 2008) and the default sensitivity parameters of the packages lme4 (Bates, Mächler, Bolker, & Walker, 2015), meta (Schwarzer, 2007), metafor (Viechtbauer, 2010), metaSEM (Cheung, 2015), and weightr (Vvea & Hedges, 1995). The aggregate data set is provided in the online supplemental material.

Results

We first begin by investigating whether and how our measures varied across the four experimental paradigms. We then present our findings on personal and situational factors impacting (dis)honest behavior. In both sections, all between-study analyses were limited to experiments that lacked normative cues (i.e., information about the [dis]honesty of other people and/or ethical reminders; limiting the dataset to $k = 508$ studies). Finally, we examine the interactions of the

experimental paradigms with the personal and situational factors by means of regression analyses. The last section turns to experiments with normative cues to thus test the moderating effect of normative cues on dishonest behavior ($k = 558$ studies).

Variation Across the Experimental Paradigms

Standardized report. Standardized reports varied strongly between and within the four experimental paradigms. As depicted in Figure 2, the standardized report ranged from $M_r = 100\%$, indicating that all participants cheated to the maximal degree, to $M_r = -33\%$, indicating that participants claimed about one third less than entirely honest participants would have done (if they observed a representative distribution of outcomes). Most experiments, however, fell somewhere between these extremes. In sender–receiver games, an average of $M_r = 51\%$ of messages were false. In coin-flip tasks, wins were reported $M_r = 31\%$ more often than would be expected from honest reporting. This value translates into an average of 66% reported wins when 50% of participants actually tossed the winning side.⁶ In die-roll tasks, the reported outcomes averaged to $M_r = 30\%$. For a six-sided die in which each reported score point (pip) paid \$1, this value converts to a mean claim of \$4.24, where \$3.50 would be expected from honest reporting.⁷ Matrix tasks yielded the smallest average standardized report of all the paradigms, indicating that $M_r = 17\%$ of all unsolved matrices were reported as solved. Overall, the standardized reports for coin-flip and die-roll tasks were relatively similar, $Q(1) = 0.27$, $p = .601$. Relative to those two experimental paradigms, sender–receiver games yielded higher standardized reports, $Q(1) = 87.75$, $p < .001$, whereas matrix tasks yielded lower standardized reports, $Q(1) = 45.04$, $p < .001$ (both comparisons combined coin-flip and die-roll tasks). Comparing the standardized report of sender–receiver games to that of the other experimental paradigms, however, makes little sense. The standardized report for sender–receiver games can never take negative values. A cleaner comparison of sender–receiver games to the other experimental paradigms is therefore the rate of liars (see next section).

Heterogeneity (in terms of I^2 and τ^2) in all experimental paradigms was large, and the standardized report did not converge to a specific value. One reason for this could be publications bias: Imprecise experiments yielded particularly extreme values of the standardized report and thus biased the grand mean of all estimates. Egger’s regression, however, suggested no such trend—except for matrix tasks, for which Eg-

⁶ For coin-flip tasks, the standardized report refers to the estimated percentage of falsely claimed wins. After a single coin toss, an honest sample would report about 50% wins, in which case $M_r = 0$. A standardized report of $M_r = 31\%$ thus translates for one-shot coin-flip tasks with a 50% honest chance of winning to 66% reported wins: $66\% = 50\%$ [estimated actually observed wins] + $(100\% - 50\%)$ [estimated actually observed losses] $\times 31\%$ [standardized report].

⁷ For die-roll tasks, the standardized report indicates the percentage of claimed score points (pips) that were over the expected average claim that would result from honest reporting. For a six-sided die in which each reported pip paid \$1, an honest sample would claim, on average, $\$3.50 = (\$1 + \$2 + \$3 + \$4 + \$5 + \$6)/6$, in which case $M_r = 0$. The standardized report of $M_r = 28\%$ thus translates to a mean claim of $\$4.24 \cong \3.50 [estimated average eligible claim] + $(6 - \$3.50)$ [estimated average maximal-possible overclaim] $\times 30\%$ [standardized report].

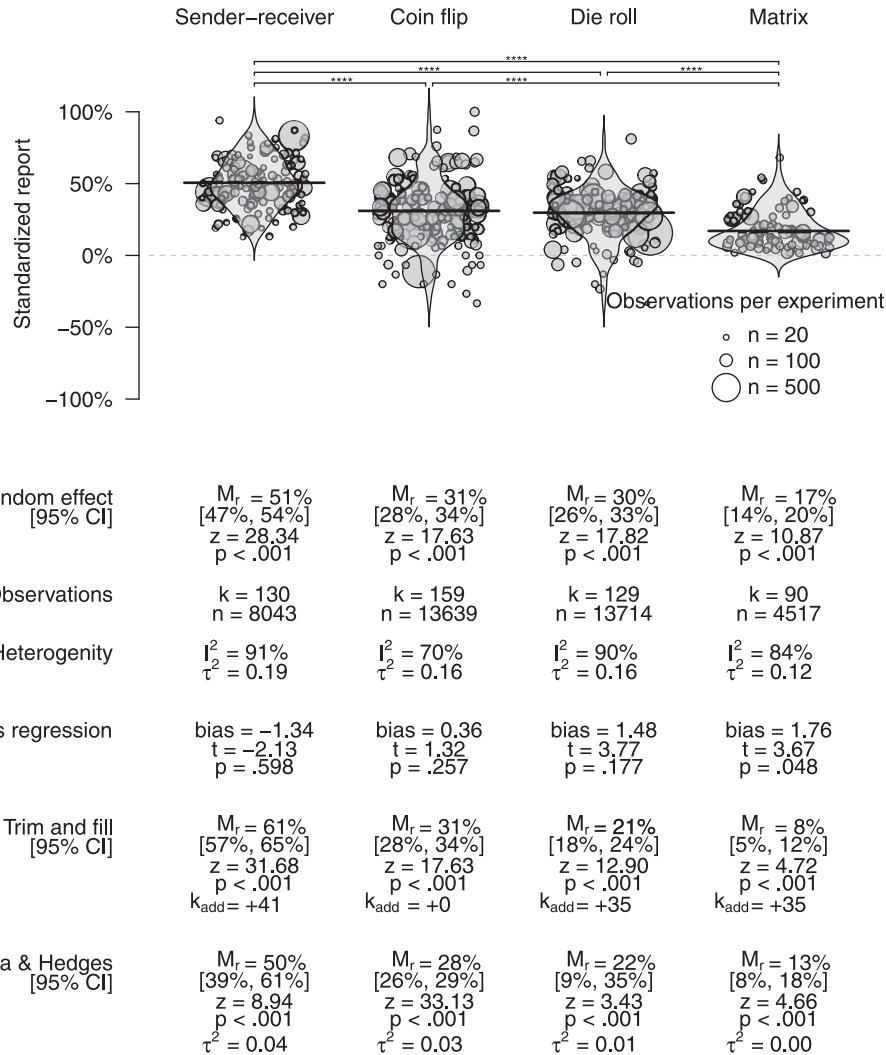


Figure 2. Violin plots showing the distribution of standardized reports by experimental paradigm. Each dot represents an experiment. The dot size indicates the number of observations. All dots are horizontally jittered for the sake of visualization. The thick, solid bars indicate the estimated mean of the standardized report. The thin, solid lines indicate the estimated mean of the standardized report after trim and fill adjustments (publication bias correction). The bars above the plots summarize the results of tests for subgroup differences, with **** $p < .0001$. The table below the figure presents the summary statistics of the random effects models: M_r is the standardized report [with 95% confidence interval]; k is the number of experiments; n is the total number of observations; I^2 is the study variance independent of the number of experiments; and τ^2 is the between-study variance. For Vevea and Hedges's weighting-function model and p -uniform we assumed cut-off points at $p = .050$. CI = Confidence interval.

ger's regression identified a marginally significant link between standard errors and standardized reports (as reported in Figure 2). Trim and fill and Vevea and Hedges's adjustments both suggested that the mean standardized report for matrix tasks should be strongly lowered. For sender-receiver games, trim and fill suggested greater mean standardized reports; for coin-flip tasks, adjustments were marginal; and for die roll-tasks, trim and fill and Vevea and Hedges's adjustments both suggested smaller mean standardized reports. Given these indications of publication bias, our subsequent analyses must be interpreted with care.

Rate of liars. We now turn to the percentage of people who behaved dishonestly, irrespective of the degree of their dishonesty. These analyses are based on a subset of experiments and observations. On average, sender-receiver games, die-roll tasks, and matrix tasks yielded relatively similar estimates of the liar rate, suggesting that 51%, 52%, and 48%, respectively, of participants lied (all pairwise comparisons: $Q[1] < 1.61$, $p > .202$; see Figure 3). By contrast, "only" about 30% of participants lied in coin-flip tasks (all pairwise comparisons: $Q[1] > 17.22$, $p < .001$). When we adjusted for publication bias, the gap in the rate of liars between coin-flip tasks and the other experimental

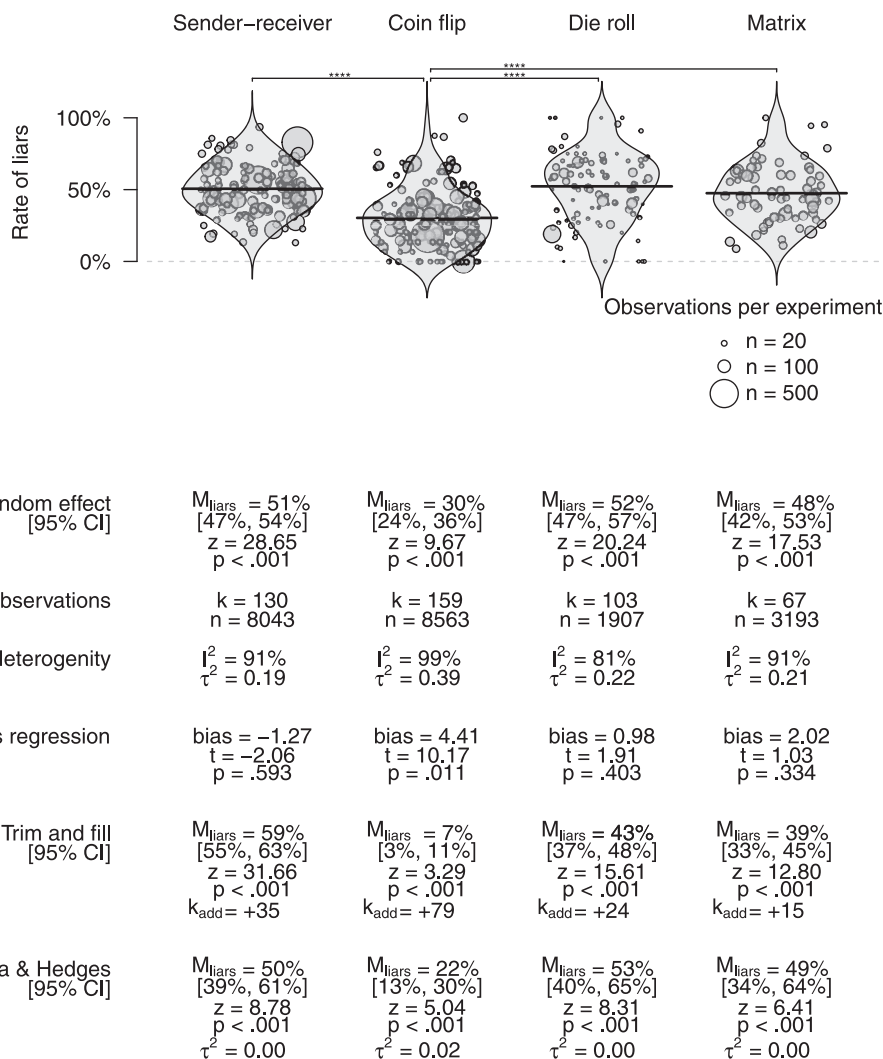
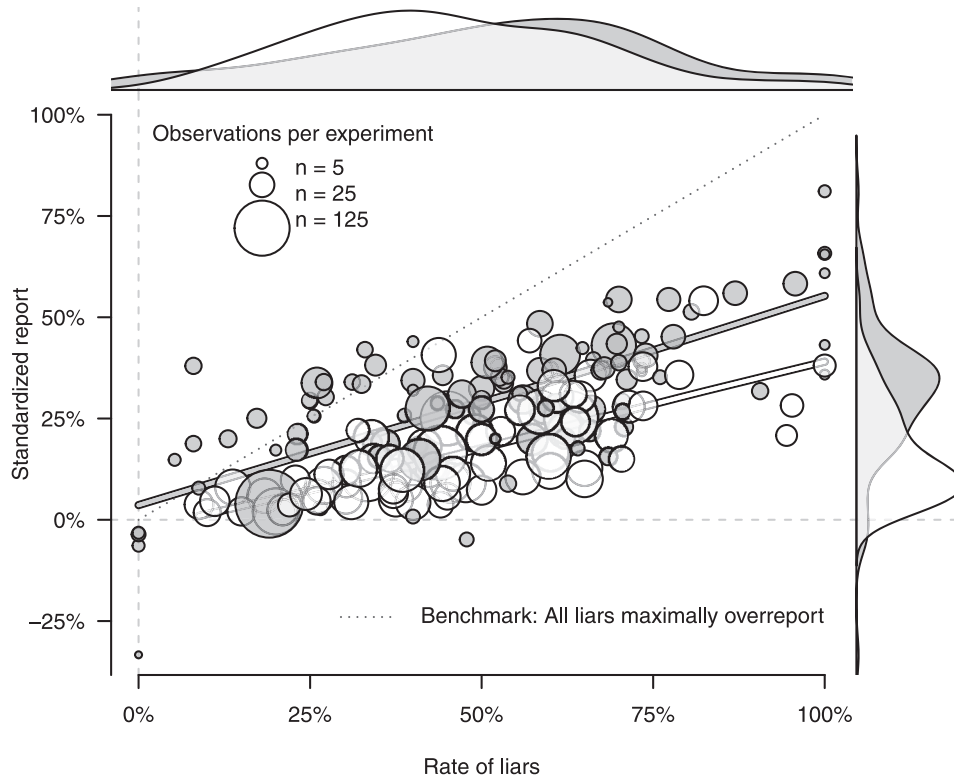


Figure 3. Violin plots showing rate of liars by experimental paradigm. The bars above the plots summarize the results of tests for subgroup differences, with **** $p < .0001$. See caption to Figure 2 for further explanation.

paradigms widened even further. Whereas Egger's regression found a link between the rate of liars and the standard errors in coin-flip tasks, no such trend was suggested for the other experimental paradigms. Trim and fill and Vevea and Hedges's adjustments both suggested lowering the mean rate of liars in coin-flip tasks.

The rate of liars and its relation to the standardized report. Die-roll and matrix tasks both have continuous outcome measures. Comparing the rate of liars with the standardized report in these two paradigms revealed a striking difference: Although a comparable rate of liars emerged for die-roll tasks ($M_{\text{liars}} = 52\%$) and matrix tasks ($M_{\text{liars}} = 48\%$), $Q(1) = 1.62$, $p = .203$, the standardized report in die-roll tasks ($M_r = 30\%$) was almost twice that in matrix tasks ($M_r = 17\%$), $Q(1) = 30.78$, $p < .001$. Taking additionally into account that M_{liars} in die-roll tasks is the upper bound (see Appendix B), this finding suggests that liars in die-roll tasks cheated to a substantially greater degree than liars in matrix tasks did (see Figure 4).

As discussed, there are several possible reasons for the difference between the two paradigms. For example, in die-roll tasks, even the most advantageous outcome is based on chance and is typically as (un)likely as the least advantageous one. In matrix tasks, in contrast, top performances and, by extension, top outcomes are typically less likely and hence less plausible than low performances and outcomes. Another reason why die-roll tasks may yield greater standardized reports than matrix tasks is the typical feature of matrix tasks to mislead participants into thinking that performance was not tracked when it was in fact traced on the individual level. Such practices could cause suspicion among the participants. To test whether the (objective) base rate of achieving the maximum outcome influences dishonesty, we calculated the chance of obtaining the highest reward through honest means for each experiment—that is, the probability of observing the highest score in die-roll tasks and the probability of solving all matrices in matrix tasks. To test the effect of individual-level measurement, we used a dummy variable indicating the analysis level (individual vs. aggregate). The results are reported in Table 2. On the



	Die roll	Matrix
Study Best fit		
Observations	k = 103 n = 1907	k = 76 n = 3484
Rate of liars	$M_{\text{liars}} = 50\%$	$M_{\text{liars}} = 45\%$
Standardized report	$M_r = 29\%$	$M_r = 16\%$
Correlation	$\rho = 0.64^{***}$	$\rho = 0.77^{***}$

Figure 4. Relationship of the rate of liars and the standardized report in experiments with continuous outcome measures. The plot in the center shows the standardized report (y axis) as a function of the rate of liars (x axis). Each dot represents a single experiment. Larger dots represent experiments with more observations. The best fitting regression lines summarize the relationship of the standardized report and the rate of liars in die-roll (gray) and matrix (white) tasks. The dotted line depicts the (hypothetical) relationship in the case that all liars claim the maximal payoff. Some die-roll tasks fell above this benchmark due to sampling errors in estimating the rate of liars (see [Appendix B](#) for details). The density plot on the right shows the distribution of the standardized report in die-roll (gray) and matrix (white) tasks. The density plot on the top shows the distribution of the rate of liars in die-roll (gray) and matrix (white) tasks. The table below the figure presents the model summary statistics. The Spearman rank correlation coefficient ρ indicates the relationship of the standardized report and the rate of liars on the level of the experiments, with $^{***} p < .001$.

one hand, 23 times more people were eligible for the highest reward in die-roll tasks than in matrix tasks ($W = 281, p < .001$; one-sided Wilcoxon's test). In 91% of the matrix tasks, not a single participant solved all matrices. Linear regression analyses showed that eligibility for the highest reward predicted the standardized report. At the same time, eligibility for the highest reward did not predict the rate of liars

thus suggesting that people shy away from claiming great rewards if such rewards are implausible (see [Table 2](#)). On the other hand, 75% of the matrix tasks misled individuals into thinking that their performance was not measured whereas "only" 4% of the die-roll tasks traced individual behavior (by making participants roll the die on a computer screen and tracing the true outcome). Such tracing was

Table 2
Eligibility for the Highest Reward and Measuring Individual Behavior Predicted Dishonest Behavior

Dependent variable	M_r	M_{liars}	M_r
Intercept	26% (2.70)***	64% (6.61)***	2% (3.52)
Percentage eligible for highest reward	39% (18.69)*	-69% (40.43)	46% (16.26)**
Individual level (dummy)	-6% (3.11)*	-36% (11.16)**	7% (4.78)
Matrix task (dummy)	-4% (3.66)	-19% (11.17)	-14% (4.64)**
Rate of liars (M_{liars})	—	—	43% (3.91)***
Observations	$k = 219$ $n = 20,451$	$k = 170$ $n = 5,100$	$k = 170$ $n = 5,100$
Residual heterogeneity	$I^2 = 84\%$ $\tau^2 = .02$	$I^2 = 85\%$ $\tau^2 = .04$	$I^2 = 58\%$ $\tau^2 = .01$
Heterogeneity accounted for	$R^2 = 33\%$	$R^2 = 10\%$	$R^2 = 72\%$

Note. Linear regression models with random effects between the experiments. Unless denoted otherwise, values refer to regression weights with standard errors in parentheses.
 * $p < .05$. ** $p < .01$. *** $p < .001$.

associated with smaller standardized reports and smaller rates of liars suggesting that people avoid lies when their behavior can be exposed on the individual level (see Table 2). The relative implausibility of high rewards in matrix tasks and the relative common use of misleading setup practices may help explain why matrix tasks yielded on average smaller standardized reports than die-roll tasks.

Personal and Situational Factors

Before assessing how personal factors are associated with dishonest behavior, we inspect the participant composition of the data set. Figure 5 shows the total number of observations per country. By far most experiments were conducted in the United States and Germany, followed primarily by other WEIRD (Western, educated, industrialized, rich, and democratic; cf. Henrich, Heine, & Norenzayan, 2010) countries. As

shown in Figure 6, there were almost as many observations for women (49%) as for men (51%). Most participants were in their 20s ($Mdn_{age} = 24$ years, $SD_{age} = 12.01$). About 77% of the experiments were conducted in samples of university and college students. As a consequence, students accounted for 63% of all observations. Of the students, 31% were majoring in business and/or economics and “only” 2% in psychology. With the exception of gender, the overall participant composition was thus far from being representative. Consequently, demographic analyses should be interpreted with care. For example, most of the data on gender effects are based on student samples and such gender effects may not generalize to other populations.

Gender. We next compared the reporting behavior of men and women in the 380 experiments in which gender information was recorded. Overall, men’s standardized reports were 4% higher than

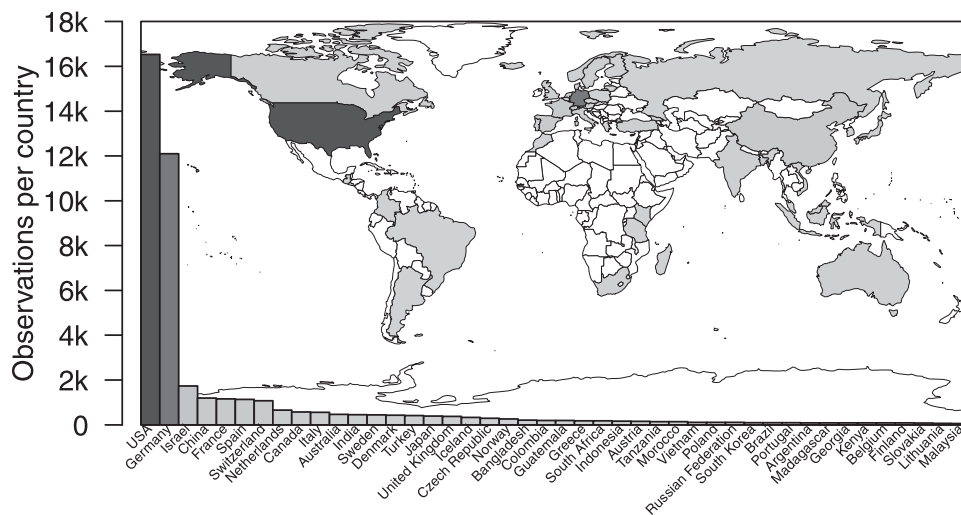


Figure 5. Observations by country: Most experiments were conducted in the United States, Germany, and other WEIRD (Western, educated, industrialized, rich, and democratic; cf. Henrich et al., 2010) countries. The shades of the bar plot represents the total number of observations. The shades are reflected in the world map and are not relative to the population sizes of the countries. Countries in white provided no observations.

This document is copyrighted by the American Psychological Association or one of its allied publishers. This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

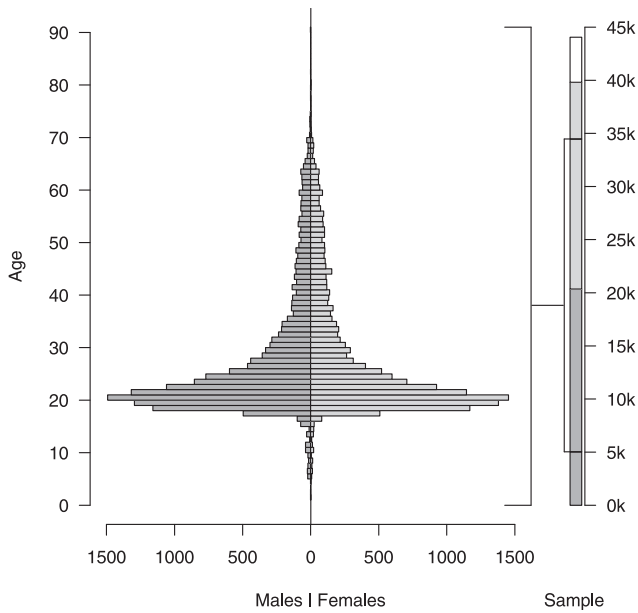


Figure 6. Observations by gender and age: There were about as many observations for men as for women; most participants were in their 20s. The demographic pyramid on the left shows the age distribution for the subsample of participants whose age and gender were known. The stacked bar plot on the right shows the total number of observations. The number of male participants is represented in dark gray. The number of female participants is shown in light gray. The number of participants whose gender was unknown is shown in white.

women's, suggesting that men behaved slightly more dishonestly than women did ($k = 373$, $n = 34,897$, $M_{\text{male } r} - M_{\text{female } r} = 33\% - 29\% = 4\%$ [rounded], 95% CI [2%, 5%], $z = 5.69$, $p < .001$). This effect was suggested within each paradigm separately (Figure 7; see also Appendix C). In total, 42% of men and 38% of women lied. This difference was statistically significant ($k = 380$, $n = 17,736$, $M_{\text{male liars}} - M_{\text{female liars}} = 42\% - 38\% = 4\%$, 95% CI [2%, 6%], $z = 4.03$, $p < .001$).

Age. To test whether age was related to dishonest behavior, we fitted a series of linear mixed effects models to the results of the 88 experiments in which age varied to at least some degree ($SD_{\text{age}} > 5$) and for which we had access to the raw data. The latter was necessary to use the participants' exact age as our independent variable. Overall, we found a small negative effect of age on several measures of dishonest behavior, suggesting that younger participants behaved more dishonestly than older participants. Every year of life lowered the standardized report by 0.28 percentage points ($k = 88$, $n = 16,495$; $b_0 = 32\%$ [$SE_0 = 3.19$], $t = 10.05$, $p < .001$; $b_{\text{age}} = -0.28\%$ [$SE_{\text{age}} = 0.05$], $t = -6.09$, $p < .001$). The age effect in the standardized report was found in all experimental paradigms but the matrix tasks (Figure 8). The rate of liars also decreased with age. Every additional year lowered the probability of lying by 0.28 percentage points ($k = 40$, $n = 7,917$; $b_0 = 53\%$ [$SE_0 = 3.06$], $t = 17.25$, $p < .001$; $b_{\text{age}} = -0.28\%$ [$SE_{\text{age}} = 0.05$], $t = -5.53$, $p < .001$; limited to experiments with individual-level data).

Investigative setting. To compare the effect of the investigative setting on dishonest behavior, we grouped all studies into

three categories—laboratory experiments, online/telephone experiments, and field experiments—and compared standardized reports between these categories (intercoder reliability: Cohen's $\kappa = 0.97$). Overall, we found no differences in standardized reports between laboratory experiments (lab: $k = 311$, $n = 21,244$, $M_r = 36\%$) and telephone/online experiments (distance: $k = 104$, $n = 14,951$, $M_r = 33\%$; lab vs. distance: $Q[1] = 0.89$, $p = .347$). However, standardized reports in field experiments were significantly lower than in laboratory experiments (field: $k = 93$, $n = 5,938$, $M_r = 26\%$; lab vs. field: $Q[1] = 10.45$, $p = .001$) and marginally lower than in telephone/online experiments (distance vs. field: $Q[1] = 3.25$, $p = .071$). Similar results were observed for the rate of liars (lab: $k = 264$, $n = 9,916$, $M_{\text{liars}} = 49\%$; distance: $k = 104$, $n = 8,333$, $M_{\text{liars}} = 41\%$; field: $k = 93$, $n = 5,938$, $M_{\text{liars}} = 26\%$; lab vs. distance: $Q[1] = 2.61$, $p = .106$; distance vs. field: $Q[1] = 4.78$, $p = .029$; lab vs. field: $Q[1] = 55.95$, $p < .001$). It is worth noting that simply comparing investigative settings is problematic, for example, because of the unequal distribution of experimental tasks per investigative setting. Coin-flip tasks were more frequently conducted in the field than sender-receiver games were. Coin-flip tasks thus contributed relatively more observations to the average standardized report of field experiments than sender-receiver games did. A more fine-grained analysis of investigative setting is summarized in Figure 9, which separates the average standardized reports by experimental paradigm. The results suggest that the standardized reports emerging from laboratory experiments were greater than those in field experiments across all experimental paradigms. To reduce the potential effect of confounds on investigative setting, for example, as a result of different participant pools, such as Amazon's MTurk, we next ran regression analyses that also considered the effect of investigative setting, controlling for other covariates.

Regression analyses allowed us to control for potential confounds in the analysis of investigative setting and to take all experiments into account, without restricting the data set to studies that provided sufficient variation within each experiment (e.g., experiments whose participants varied in age or that assessed gender). Table 3 summarizes two regression models that predict the measures of dishonest behavior (M_r , M_{liars}) by lumping together the experimental paradigms and adding dummy variables for the experimental paradigms. Table 4 presents additional regression models that predict the standardized reports in each of the four experimental paradigms.

Online/telephone experiments. Overall, there was as much (dis)honest behavior in lab experiments as in experiments conducted online or via telephone (see Table 3). The exception was coin-flip tasks, in which the standardized reports were on average 24 percentage points lower in online/telephone experiments than in lab studies (see Table 4).

Field experiments. Dishonesty was more prevalent in lab experiments than in field experiments. The standardized reports were 13 percentage points lower in field settings and the rate of liars was 15 percentage points lower (see Table 3). This effect was independently observed in sender-receiver games and in coin-flip tasks; here, standardized reports were 14 and 25 percentage points lower, respectively, in field experiments than in lab studies (see Table 4).

Nonstudents. The results of experiments with nonstudent groups ($\kappa = 1.00$) were largely similar to those with students who

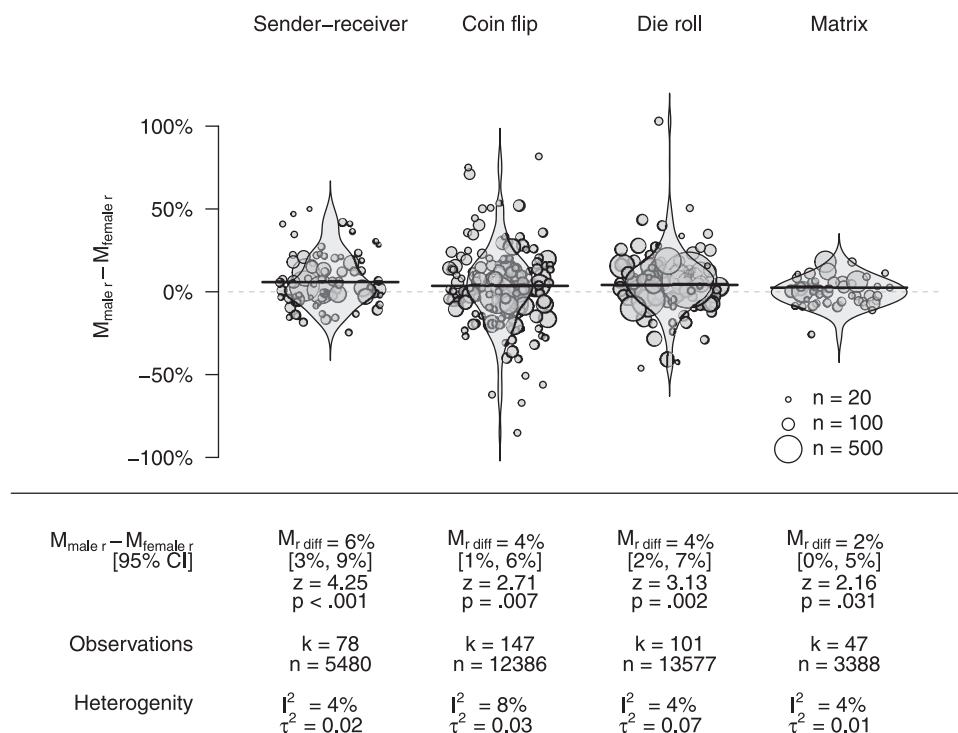


Figure 7. Violin plots showing gender differences in the standardized report across the four paradigms. See caption to Figure 2 for further explanation.

majored in disciplines other than economics (see Table 3). The only exception was sender–receiver games, in which the standardized reports for nonstudent samples were 18 percentage points higher than those for noneconomic student samples (see Table 4), suggesting that students who majored in a discipline other than economics behaved more honestly than nonstudent population groups.

Economics students. Within the student population, we tested whether majoring in economics was associated with more or less dishonest behavior. The percentage of economics students in the sample ($\kappa = 0.93$) did not predict any of the outcome measures (Table 3; Table 4).

MTurk. Experiments conducted on Amazon’s MTurk ($\kappa = 1.00$) yielded an average 11 percentage points more liars compared with experiments conducted with student populations (see Table 3). The effect was independently found in sender–receiver games and coin-flip tasks (26 and 21 percentage points greater standardized reports, respectively; Table 4).

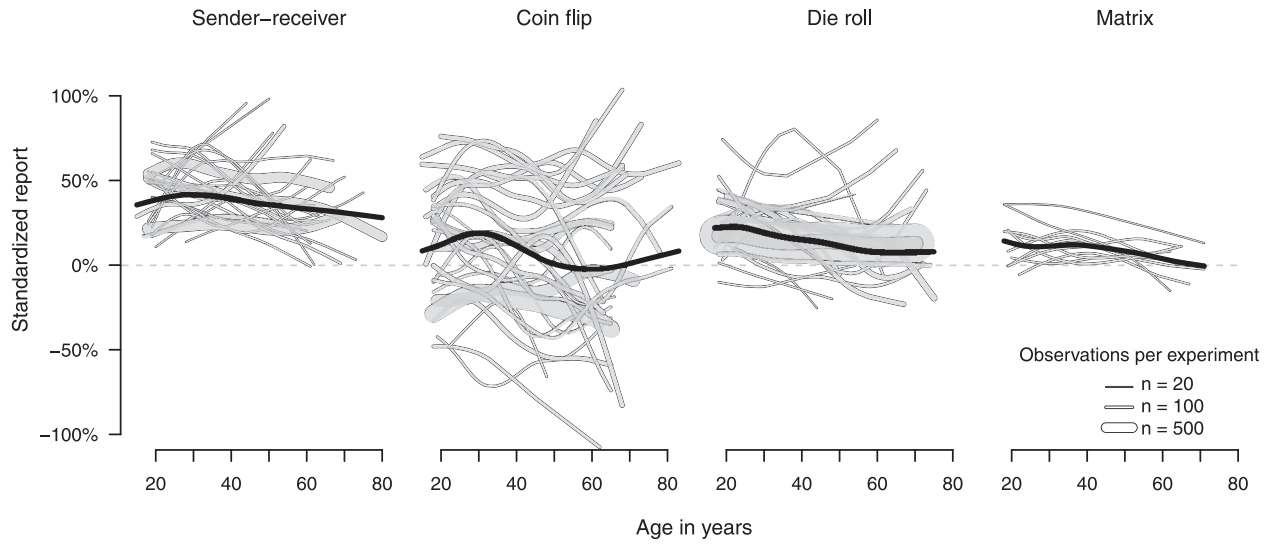
Normative cues. To test the effect of normative cues we calculated a variable that could range between -1 and $+1$. A value of -1 indicates that participants were informed that all other participants cheated, 0 indicates the absence of information on other participants and the absence of ethical reminders, and $+1$ indicates the presence of an ethical reminder ($\kappa = 0.93$). Overall, the standardized report and the rate of liars were both strongly affected by normative cues (see Table 3). This effect, however, was largely due to sender–receiver games (see Table 4).

Experimental deception. Experiments in which participants were intentionally misinformed ($\kappa = 0.95$) had eight percentage

points smaller standardized reports and 7 percentage points smaller rates of liars (see Table 3). The effect was largely attributable to sender–receiver games, for which experimental deception always meant misinforming participants about the existence of a receiver (and possibly other forms of experimental deception too). In such experiments, participants sent 12 percentage points fewer false messages (see Table 4).

Maximal externality. To assess the effect of externalities, we calculated the upper limit of harm that participants could inflict on other participants by behaving dishonestly ($\kappa = 0.96$). In sender–receiver games, receivers typically gain less when they trust a false message, although the amount lost can vary (e.g., Gneezy, 2005). The other three paradigms do not typically involve such externalities. However, some experiments did change the usual setup, for example, by making one participant’s gain another participant’s loss (constant sum; e.g., Gino et al., 2013). Having calculated the maximal externality that could be inflicted on other participants, we converted these costs from national currency units to U.S. dollars using 2015 purchasing power parity exchange rates (see Appendix B for details). The results suggest that the effect of maximal externality is limited to sender–receiver games (Table 3; Table 4). Here, every additional U.S. dollar kept from receivers was associated with a 1 percentage point decrease in false messages sent. Notably, the effect of maximal externality was independent of what the participants themselves stood to gain, as discussed next.

Maximal gain. To test whether reward sizes affected dishonesty, we calculated the upper limit that participants could themselves gain from behaving dishonestly ($\kappa = 0.96$), again converting national currency units to 2015 U.S. dollar purchasing power parity. In sender–



	Sender-receiver	Coin flip	Die roll	Matrix
Intercept	$b_0 = 0.48^{****}$ (0.04)	$b_0 = 0.21^{**}$ (0.07)	$b_0 = 0.31^{****}$ (0.04)	$b_0 = 0.18^{****}$ (0.04)
Age	$b_{age} = -0.23\%^{**}$ (0.08)	$b_{age} = -0.28\%^{**}$ (0.09)	$b_{age} = -0.28\%^{****}$ (0.05)	$b_{age} = -0.19\%$ (0.10)
Observations	k = 26 n = 2912	k = 29 n = 6251	k = 23 n = 6805	k = 10 n = 462

Figure 8. Age effects in standardized reports across the four paradigms (for experiments with participants whose age varied by at least five standard deviations). The graphs depict the standardized report per experiment (y axis) as a function of participants' age (x axis). The lines were smoothed by local polynomial fitting and weighted by the number of observations. The line width indicates the total number of observations per experiment. The black lines are the estimated means of the standardized report per paradigm. The table below the figures reports the summary statistics of mixed linear regression models with random intercepts between experiments with standard errors in parentheses, and with **** $p < .0001$. ** $p < .01$.

receiver games and in coin-flip tasks, greater potential reward sizes were linked to higher standardized reports (see Table 4). Specifically, every additional U.S. dollar increased the percentage of false messages sent by one percentage point in sender receiver games. A similar effect was observed in coin-flip tasks, in which every additional U.S. dollar in maximal gain increased the standardized report by 0.46 percentage points.

We further inspected the effect of reward size on dishonesty in a more fine-grained analysis. A total of 33 studies experimentally manipulated the incentive size and randomly assigned participants to either a high or a low gain condition while keeping all other parameters constant. This setup permits to isolate the effect of an increase in gains on dishonest behavior. To compare the effect of maximum gain (MG) across the experiments, we subtracted the standardized report of the highest paying condition in a given experiment (M_r for high MG) from the standardized report of the lowest paying condition (M_r for low MG). We thus obtained the mean differences, D , in the standardized report for every experiment that manipulated reward size, $D (= M_r \text{ for high MG} - M_r \text{ for low MG})$. We then combined all mean differences through a random effects model. In essence, the model estimates to what degree an increase in reward size causes the standardized report to change. Overall, the results confirm that greater reward sizes were associated with higher standardized reports in sender-receiver games. No such effect, however, was observed for the other three paradigms, including coin-flip tasks (see Figure 10).

The effect of maximal gains on (dis)honest behavior was potentially confounded by the reward size being increased to a greater extent in sender-receiver games than in the other paradigms. To further inspect the effect of MG, we calculated by how many times the maximal gain in the highest paying condition (MG_{high}) was increased relative to the maximal gain in the lowest paying condition (MG_{low})—the relative increase, $MG_{factor} (= MG_{high}/MG_{low})$. In addition, we calculated the absolute difference in the maximum amount paid in the highest paying condition relative to the lowest paying condition, $MG_{difference} (= MG_{high} - MG_{low})$. We then independently regressed MG_{factor} and $MG_{difference}$ to predict the mean difference in the standardized report, D (see Figure 11). Neither a relative increase in maximal gains, MG_{factor} , nor an absolute increase, $MG_{difference}$, predicted the standardized report. Consequently, the observed effect of increasing reward sizes on standardized reports in sender-receiver games seems to have been caused by factors other than just its magnitude.

Discussion

Various experimental paradigms have been used to assess dishonest behavior. Our first goal was to synthesize four of the most widely used paradigms. Overall, the results suggest that the degree of (dis)honest behavior hinge on the experimental paradigm. For

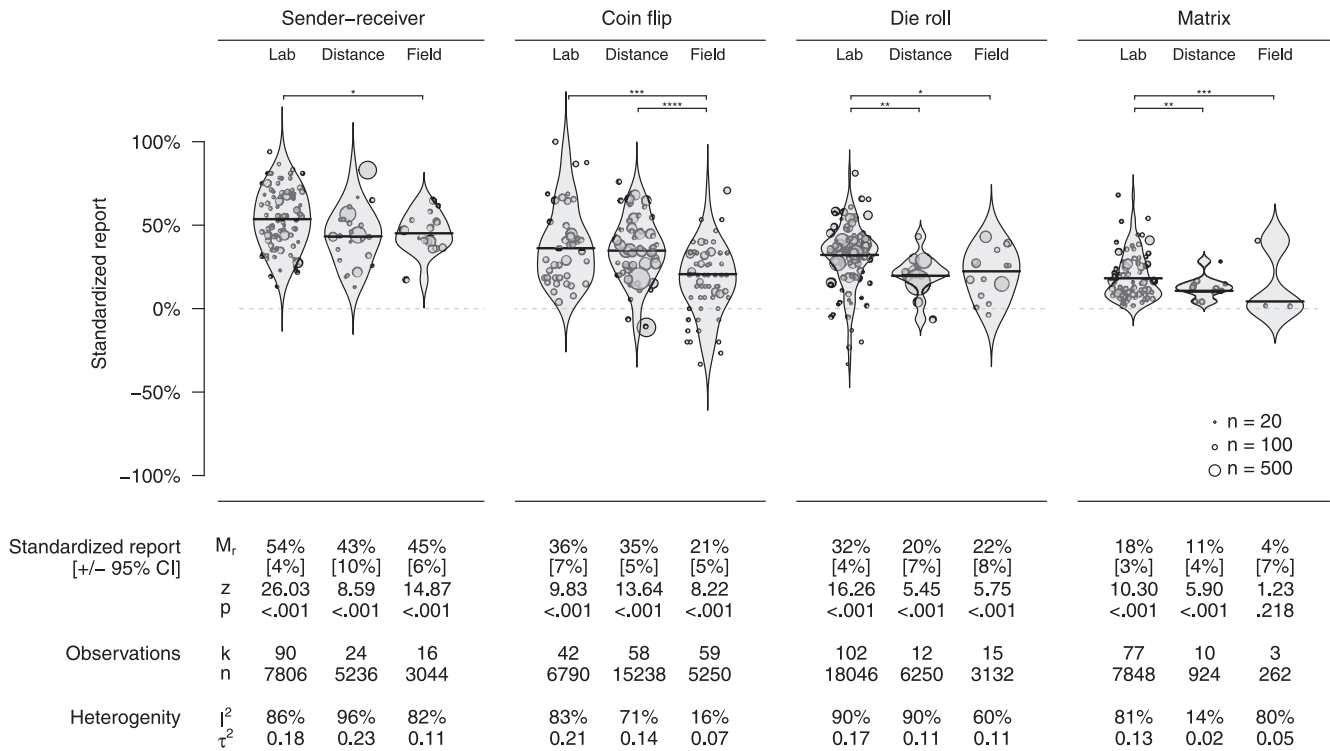


Figure 9. Violin plots showing the distribution of standardized reports by experimental paradigm and investigative setting. The bars above the plots summarize the results of tests for subgroup differences, with * $p < .05$, ** $p < .01$, *** $p < .001$, and **** $p < .0001$. See caption to Figure 2 for further explanation.

example, liars in die-roll tasks cheated substantially more than liars in matrix tasks. Potentially, this difference could be explained with the skill component unique to the matrix tasks (e.g., people might feel more uncomfortable with overreporting their skill-based results than chance-based results). But differences in dishonest behavior were observed even between the two most similar experimental paradigms: Coin-flip and die-roll tasks. The most apparent difference between the two is their outcome measure, which is either dichotomous (coin-flip tasks) or continuous (die-roll tasks). Many reviews consequently have treated them as two versions of the same experimental paradigm (e.g., Abeler et al., 2016; Jacobsen et al., 2018; Rosenbaum et al., 2014). Yet our results suggest that the two paradigms are associated with notable differences in the rate of liars. In addition, some of our analyses of situational and personal factors led to different conclusions for coin-flip tasks versus die-roll tasks. Even the relatively similar standardized reports for coin-flip and die-roll tasks (see Figure 2) seemed confounded by differences in the experimental setup practices (see Table 3). The dissimilarity of the two paradigms is also reflected in two recent large-scale analyses that used either die-roll or coin-flip tasks to compare the reporting behavior of citizens of several countries. Whereas the corruption index on the country level predicted reporting behavior in die-roll tasks (Gächter & Schulz, 2016), there was no such effect in coin-flip tasks (Pascual-Ezama et al., 2015). We also found that there was substantial evidence for publication bias in most measures of dishonesty and with large differences between the paradigms. For example, cor-

recting for publication bias largely increased the standardized report in sender-receiver games, but it decreased the standardized report in die-roll tasks.

Next to heterogeneity across paradigms, we also observed substantial heterogeneity *within* all four experimental paradigms, suggesting that the manifestation of dishonest behavior strongly depends on situational and personal factors. In fact, assessing these factors and their interaction with experimental paradigm represented our second and third goal, respectively. In brief, we identified the following personal and situational factors: Overall, males behaved slightly more dishonestly than females. Younger participants behaved somewhat more dishonestly than older participants (except in matrix tasks). Nonstudents behaved largely as (dis)honestly as noneconomics students (except in sender-receiver games, where nonstudents behaved more dishonestly). And economics students behaved largely as dishonestly as noneconomics students in all experimental paradigms. MTurk workers lied more than student samples (but only in sender-receiver games and coin-flip tasks). Online/telephone experiments and laboratory experiments largely yielded similar estimates of dishonesty (except in coin-flip tasks, in which more dishonesty was observed when conducted in the laboratory). Field experiments were associated with less dishonesty than laboratory experiments (this effect was largely due to sender-receiver games and coin-flip tasks).

We also found that an experimental practice, frequently used in psychology but virtually prohibited in experimental economics, affected the rate of dishonesty (Hertwig & Ortmann, 2001). Spe-

Table 3
Predictors of Different Measures of Dishonest Behavior Across the Four Paradigms

Dependent variable (reference category)	M_r	M_{liars}
Intercept	35% (3.78) [†]	38% (4.65) [†]
Experimental paradigm (coin-flip task)		
Sender–receiver	17% (2.44) [†]	16% (2.81) [†]
Die roll	–5% (2.55) [*]	17% (3.18) [†]
Matrix	–17% (2.86) [†]	11% (3.59) ^{**}
Investigative setting (laboratory)		
Online/telephone	–6% (3.75)	–7% (4.60)
Field experiment	–13% (2.53) [†]	–15% (2.84) [†]
Participant sample (noneconomics students)		
Nonstudents	3% (3.29)	0% (4.09)
Economics students	0% (2.62)	4% (3.20)
Mechanical Turk workers	5% (3.86)	11% (4.59) [*]
Normative cue	–16% (2.99) [†]	–18% (3.57) ^{****}
Experimental deception	–8% (2.21) ^{****}	–7% (2.82) [*]
Maximal externality	–0% (.07)	–0% (.13)
Maximal gain	–0% (.06)	–0% (.13)
Observations	$k = 558$ $n = 44,050$	$k = 507$ $n = 23,332$
Residual heterogeneity	$I^2 = 84\%$ $\tau^2 = .02$	$I^2 = 91\%$ $\tau^2 = .04$
Heterogeneity accounted for	$R^2 = 54\%$	$R^2 = 58\%$

Note. Linear regression models with random effects at the experiment level. Unless denoted otherwise, values refer to beta weights with standard errors in parentheses.

[†] $p < .0001$. * $p < .05$. ** $p < .01$. **** $p < .001$.

cifically, deceiving participants about the procedure led to less dishonesty (at least in matrix tasks, Table 2, and sender–receiver games, Table 4). Possibly this is because experimental deception raised suspicion among the participants. As outlined before, various social sciences theories suggest that recognizing (or suspect-

ing) experimental deception might be interpreted as a signal that breaching of social norms is legitimate and consequently dishonest behavior should become more prevalent. Our results, however, suggest that recognized experimental deception has the opposite effect: It appears to heighten the sense among participants that socially undesirable behaviors (e.g., dishonesty) are being observed (e.g., secretly recorded), thus possibly prompting participants to behaviors that makes them look good in the eyes of observers (Hertwig & Ortmann, 2008a). This explanation is in line with the results of an early experiment that found participants who experienced experimental deception provided a more favorable self-presentation (Silverman, Shulman, & Wiesenthal, 1970; see also: Hertwig & Ortmann, 2008b; Ortmann & Hertwig, 2002). Whatever the correct explanation is, these results suggest that experimental deception in the case of the study of dishonest behavior may systematically interact with the phenomenon to be studied. Therefore, researchers investigating dishonesty should be wary of using this tool.

Finally, payoffs appear to matter. Decreasing externalities and increasing reward sizes were associated with more dishonesty in sender–receiver games—and to some degree in coin-flip tasks. Further analyses suggested that the effect of rewards on dishonesty could neither be linked to a relative nor an absolute increase in reward sizes and that it was only found in sender–receiver games.

Limitations

Some limitations of our analysis warrant consideration. First, the quantitative estimates have to be interpreted against the background of potential biases. We found substantial indication of publication bias in almost all measures of dishonest behavior. Several “correction” methods indicated that the magnitude of dishonest behavior may be falsely estimated (see also Appendix C). Moreover, the participant composition was far from representative of the general population. For example, student samples contributed substantially more data points to the analyses than

Table 4
Predictors of the Standardized Report by Experimental Paradigm

Paradigm (reference category)	Sender–Receiver	Coin flip	Die roll	Matrix
Intercept	41% (5.27) [†]	62% (8.78) [†]	16% (8.06) [*]	25% (12.92)
Investigative setting (laboratory)				
Online/telephone	–11% (6.05)	–24% (8.23) ^{**}	–9% (10.20)	–13% (13.50)
Field experiment	–14% (4.53) ^{**}	–25% (5.25) [†]	2% (8.12)	–11% (7.77)
Participant characteristics (noneconomics students)				
Nonstudents	18% (5.55) ^{**}	–17% (7.61)	16% (8.13)	–5% (13.16)
Economics students	–3% (3.95)	–16% (9.78)	–9% (6.83)	5% (6.53)
Mechanical Turk workers	26% (6.89) ^{****}	21% (10.88) [*]	11% (11.79)	—
Normative cue	–20% (3.47) [†]	–18% (15.11)	—	–10% (5.94)
Experimental deception	–12% (3.44) ^{****}	–35% (18.61)	–3% (6.57)	–3% (3.19)
Maximal externality	–1% (.45) [*]	0% (.17)	0% (.21)	0% (.24)
Maximal gain	1% (.45) ^{**}	0% (.17) ^{**}	0% (.20)	0% (.05)
Observations	$k = 165$ $n = 9,243$	$k = 163$ $n = 13,879$	$k = 129$ $n = 14,835$	$k = 101$ $n = 6,093$
Residual heterogeneity	$I^2 = 84\%$ $\tau^2 = .02$	$I^2 = 64\%$ $\tau^2 = .02$	$I^2 = 89\%$ $\tau^2 = .03$	$I^2 = 77\%$ $\tau^2 = .01$
Heterogeneity accounted for	$R^2 = 37\%$	$R^2 = 13\%$	$R^2 = 0\%$	$R^2 = 20\%$

Note. Linear regression models with random effects at the experiment level. Unless denoted otherwise, values refer to beta weights with standard errors in parentheses.

[†] $p < .0001$. * $p < .05$. ** $p < .01$. **** $p < .001$.

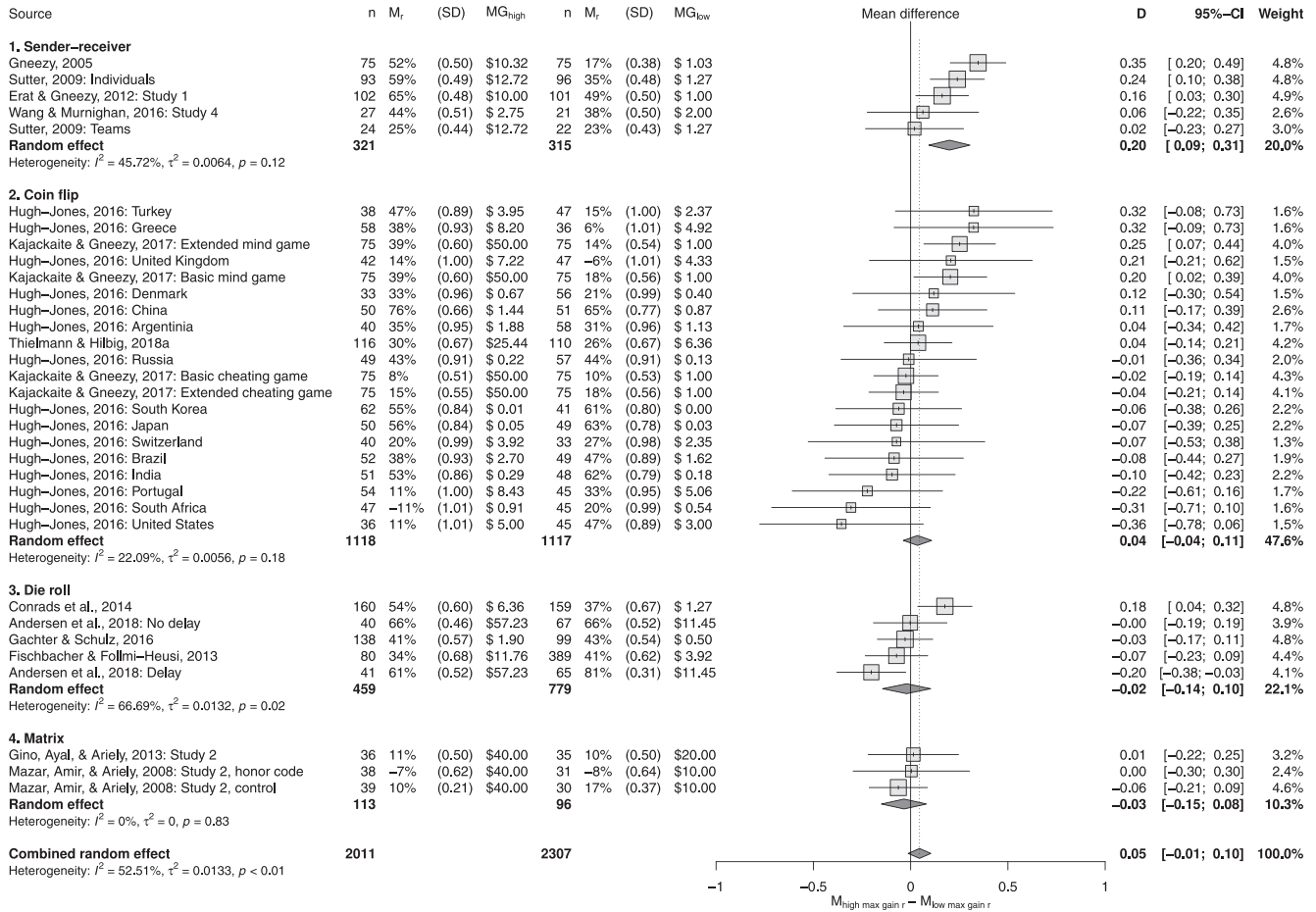


Figure 10. Forest plot: Increasing the incentive was associated with higher standardized reports in sender–receiver games but not in the other experimental paradigms.

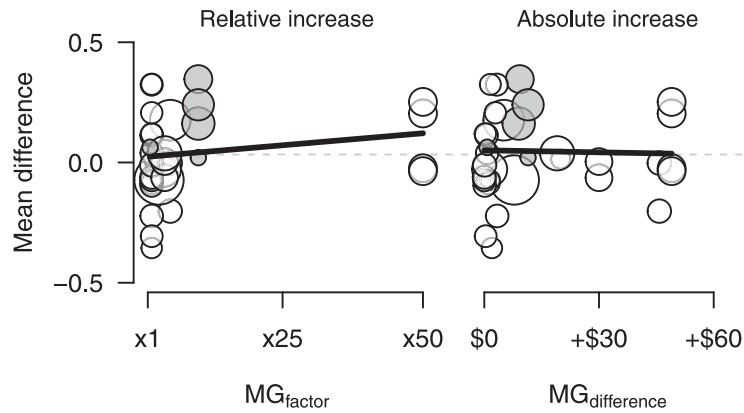
other population groups. However, the gender effect and potentially all other effects may have been solely due to student populations. To what degree these effects generalize beyond the typical student population therefore remains an open, empirical question (see, e.g., Abeler et al., 2014). Overall, more representative participant pools seem highly desirable for future research on dishonesty.

A second limitation is that our analyses of personal and situational factors were cross-sectional and correlational. As such, these analyses do not permit causal interpretation. For example, age effects, which were assessed cross-sectionally may have been confounded by cohort effects. Only longitudinal analyses can disentangle how aging as opposed to shared experiences of events (e.g., collective experience of societal corruption) affects behavior (see also Gächter & Schulz, 2016). We are not aware of any such longitudinal analyses on dishonest behavior. Third, some of our regression models accounted for relatively little heterogeneity, suggesting a poor fit of the covariates to the data. In the regression models on die-roll and matrix tasks (see Table 4), in particular, variables making important contributions to the relatively large heterogeneity seemed to be lacking. For matrix tasks, one reason may be our central measure of dishonest behavior, the standardized report. Matrix tasks typically report dishonest behavior in terms of

the absolute number of unsolved matrices claimed as solved—not as the percentage thereof (however, see Appendix C). For sender–receiver games, specific idiosyncrasies in the experimental setup may explain part of the large variance in the amount of dishonest behavior observed (see Appendix C).

Directions for Future Research

Much of the correlational evidence seems worth exploring through more rigorous experimental techniques. For example, most dishonest behavior seemed relatively robust to increases in reward size—except in sender–receiver games. Why are people corruptible in sender–receiver games but not in other experimental paradigms? More generally, findings on the rates of liars and levels of (dis)honest behavior were mixed between the experimental paradigms. Given that all four experimental paradigms are widely used to examine dishonest behavior, it is important to find out why they yielded such disparate results. For example, our findings suggest that liars overreport more in die-roll tasks than they do in matrix tasks. Generalizing from either experimental paradigm would likely imply different conclusions about human nature in general and the nature of dishonest behavior specifically.



Intercept	$b_0 = 0.02$ (0.04)	$b_0 = 0.05$ (0.04)
Maximal gain (MG)	$b_{MG_{factor}} = 0.20\%$ (0.17)	$b_{MG_{difference}} = -0.03\%$ (0.16)
Observations	$k = 33$ $n = 4456$	$k = 33$ $n = 4456$
Residual heterogeneity	$I^2 = 53\%$ $\tau^2 = 0.01$	$I^2 = 53\%$ $\tau^2 = 0.01$
Heterogeneity accounted for	$R^2 = 0\%$	$R^2 = 0\%$

Figure 11. Neither a relative increase in maximal gains nor an absolute increase in maximal gains predicted change in the standardized report. Dots in gray represent sender–receiver games; dots in white represent either coin-flip tasks, die-roll tasks, or matrix tasks. The left panel shows change in the standardized report, D , as a function of by how many times the reward size was increased, MG_{factor} . The right panel shows the change in the standardized report, D , as a function of the absolute difference in maximal gains, $MG_{difference}$. Best-fit regressions are shown by solid black lines. The table below the figures describes the summary statistics of linear regression models with beta weights and standard errors in parentheses (all $ps > .050$).

The desire to appear honest may explain some of the differences in observations between the experimental paradigms. However, experimental manipulations would be required to infer causality about when and why people shy away from overreporting. Such experiments could also help determine whether participants want to appear honest to themselves (producing guilt when not following some internalized norm; as suggested by Mazar et al., 2008) or to others (producing shame when not following some social norm; as implied by the results of Abeler et al., 2014; Conrads & Lotz, 2015; Gneezy et al., 2018; Yaniv & Siniver, 2016). For example, future experiments may manipulate the degree to which experimenters are blind to the payment (Hoffmann, Diedenhofen, Verschuere, & Musch, 2015; Mazar et al., 2008), thus permitting exogenous control for the (perceived) likelihood of appearing dishonest to the experimenter.

An important issue for the future is to better understand why different experimental paradigms result in notably different behaviors—for example, the reporting behavior of similar participant groups in die-roll tasks versus matrix tasks (Gino et al., 2013; Gravert, 2013; Kajackaite, 2018). Different internalized and/or social norms may exist for the reporting of outcomes that are

randomly generated (die-roll tasks) than for outcomes indicative of one's own effort (matrix tasks). For instance, standards of academic integrity in colleges and universities emphasize issues such as cheating, plagiarism, unfair advantages, and misrepresentation. It could well be the case that skill-based tasks represented as achievement tasks are more likely to remind participants of a code of academic integrity than are chance-based tasks (see Mazar et al., 2008). Investigations of the possible impact of specific domains and possibly more domain-specific norms is likely to also shed light on the prevalence and degree of honest and dishonest behaviors outside of the laboratory.

Our finding of different results emerging across experimental designs relates to a methodological and theoretical concern that Egon Brunswik (1943, 1944) described as the double standard in the practice of sampling in psychological research: Why is the logic of generalization from participant population to the general population not equally employed to the sampling of experimental tasks? Brunswik's proposed solution was representative design (as opposed to systematic design), which involves randomly sampling experimental stimuli from the environment or designing stimuli in which theoretically or practically important properties are pre-

served (Dhimi, Hertwig, & Hoffrage, 2004). It seems a promising approach to put more conceptual thought into what structural properties exactly differ between real-world situations that offer the opportunity to behave dishonestly and how those situations, their properties, and their prevalence can be represented in the experimental microworlds.

Conclusion

In July 2017, Luiz Inácio Lula da Silva, the former president of Brazil, was sentenced to nearly 10 years in prison for corruption and money laundering (“Luiz Inácio Lula da Silva. The Rise and fall of a Brazilian leader,” 2017). Just a few days earlier, the former Israeli prime minister Ehud Olmert was granted parole on charges of bribery, fraud, obstruction of justice, and breach of trust (“Ex-PM Olmert released,” 2017). In 2018, the former French president Nicolas Sarkozy has been facing charges including corruption and illegal campaign financing (“Nicolas Sarkozy to face trial for corruption and influence peddling,” 2018). These are just three recent examples from a long list of prominent figures who have been found guilty of deceiving the public. Exploring the psychological foundations of unethical behavior in the personal, professional, and political spheres is clearly as important as ever. But this also means carefully scrutinizing and understanding the experimental tools used to this end. Our statistical synthesis of 558 experiments offers answers to many of the ongoing debates on who behaves dishonestly and under what circumstances. We showed that the degree and direction of (dis)honest behavior depends on the experimental paradigms, on situational factors (e.g., investigative setting, reward size, externalities), and on personal factors (e.g., gender, age). Yet many questions remain (e.g., why the reward size matters for sender–receiver games but not for the other experimental paradigms?). We hope that this metaanalytical synthesis will help researchers to identify crucial questions about the psychology of honest and dishonest behaviors.

References

- Abeler, J., Becker, A., & Falk, A. (2014). Representative evidence on lying costs. *Journal of Public Economics*, *113*, 96–104. <http://dx.doi.org/10.1016/j.jpubeco.2014.01.005>
- Abeler, J., Nosenzo, D., & Raymond, C. (2016, November). *Preferences for truth-telling* (CESifo Working Paper No. 6087). Retrieved from https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2866381
- Ackert, L. F., Church, B. K., Kuang, X., & Qi, L. (2011). Lying: An experimental investigation of the role of situational factors. *Business Ethics Quarterly*, *21*, 605–632. <http://dx.doi.org/10.5840/beq201121438>
- Andersen, S., Gneezy, U., Kajackaite, A., & Marx, J. (2018). Allowing for reflection time does not change behavior in dictator and cheating games. *Journal of Economic Behavior & Organization*, *145*, 24–33. <http://dx.doi.org/10.1016/j.jebo.2017.10.012>
- Anderson, C. A., & Bushman, B. J. (2002). Human aggression. *Annual Review of Psychology*, *53*, 27–51. <http://dx.doi.org/10.1146/annurev.psych.53.100901.135231>
- Angelova, V., & Regner, T. (2013). Do voluntary payments to advisors improve the quality of financial advice? An experimental deception game. *Journal of Economic Behavior & Organization*, *93*, 205–218. <http://dx.doi.org/10.1016/j.jebo.2013.03.022>
- Aoki, K., Akai, K., & Onoshiro, K. (2013, April). *An apology for lying* (ISER Discussion Paper). Retrieved from <http://econpapers.repec.org/paper/dprwpaper/0786.htm>
- Arbel, Y., Bar-El, R., Siniver, E., & Tobol, Y. (2014). Roll a die and tell a lie: What affects honesty? *Journal of Economic Behavior & Organization*, *107*, 153–172. <http://dx.doi.org/10.1016/j.jebo.2014.08.009>
- Aydogan, G., Jobst, A., D’Ardenne, K., Müller, N., & Kocher, M. G. (2017). The detrimental effects of oxytocin-induced conformity on dishonesty in competition. *Psychological Science*, *28*, 751–759. <http://dx.doi.org/10.1177/0956797617695100>
- Bassarak, C., Leib, M., Mischkowski, D., Strang, S., Glöckner, A., & Shalvi, S. (2017). What provides justification for cheating—Producing or observing counterfactuals? *Journal of Behavioral Decision Making*, *30*, 964–975. <http://dx.doi.org/10.1002/bdm.2013>
- Bates, D., Mächler, M., Bolker, B. M., & Walker, S. C. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*. Advance online publication. <http://dx.doi.org/10.18637/jss.v067.i01>
- Batson, C. D., Kobrynowicz, D., Dinnerstein, J. L., Kampf, H. C., & Wilson, A. D. (1997). In a very different voice: Unmasking moral hypocrisy. *Journal of Personality and Social Psychology*, *72*, 1335–1348. <http://dx.doi.org/10.1037/0022-3514.72.6.1335>
- Becker, G. S. (1968). Crime and punishment: An economic approach. *Journal of Political Economy*, *76*, 169–217. <http://dx.doi.org/10.1086/259394>
- Bicchieri, C. (2006). *The grammar of society: The nature and dynamics of social norms*. United Kingdom: Cambridge University Press.
- Biziou-van-Pol, L., Haenen, J., Novaro, A., Liberman, A. O., & Capraro, V. (2015). Does telling white lies signal pro-social preferences? *Judgment and Decision Making*, *10*, 538–548.
- Bonetti, S. (1998). Experimental economics and deception. *Journal of Economic Psychology*, *19*, 377–395. [http://dx.doi.org/10.1016/S0167-4870\(98\)00012-9](http://dx.doi.org/10.1016/S0167-4870(98)00012-9)
- Brunswik, E. (1943). Organismic achievement and environmental probability. *Psychological Review*, *50*, 255–272. <http://dx.doi.org/10.1037/h0060889>
- Brunswik, E. (1944). Distal focussing of perception: Size-constancy in a representative sample of situations. *Psychological Monographs*, *56*, 1–49. <http://dx.doi.org/10.1037/h0093505>
- Buccioli, A., & Piovesan, M. (2011). Luck or cheating? A field experiment on honesty with children. *Journal of Economic Psychology*, *32*, 73–78. <http://dx.doi.org/10.1016/j.joep.2010.12.001>
- Burks, S. V., & Krupka, E. L. (2012). A multimethod approach to identifying norms and normative expectations within a corporate hierarchy. Evidence from the financial services industry. *Management Science*, *58*, 203–217. <http://dx.doi.org/10.1287/mnsc.1110.1478>
- Cadsby, C. B., Du, N., & Song, F. (2016). In-group favoritism and moral decision-making. *Journal of Economic Behavior & Organization*, *128*, 59–71. <http://dx.doi.org/10.1016/j.jebo.2016.05.008>
- Cagala, T., Glogowsky, U., & Rincke, J. (2018). *Commitment to a no-cheating rule can increase cheating*. <http://dx.doi.org/10.2139/ssrn.3111855>
- Cai, W., Huang, X., Wu, S., & Kou, Y. (2015). Dishonest behavior is not affected by an image of watching eyes. *Evolution and Human Behavior*, *36*, 110–116. <http://dx.doi.org/10.1016/j.evolhumbehav.2014.09.007>
- Cappelen, A. W., Sørensen, E. Ø., & Tungodden, B. (2013). When do we lie? *Journal of Economic Behavior & Organization*, *93*, 258–265. <http://dx.doi.org/10.1016/j.jebo.2013.03.037>
- Capraro, V. (2017). Does the truth come naturally? Time pressure increases honesty in one-shot deception games. *Economics Letters*, *158*, 54–57. <http://dx.doi.org/10.1016/j.econlet.2017.06.015>
- Capraro, V. (2018). Gender differences in lying in sender–receiver games: A meta-analysis. *Judgment and Decision Making*, *13*, 345–355. Retrieved from <http://journal.sjdm.org/18/18619a/jdm18619a.pdf>
- Charness, G., Blanco, C., Ezquerra, L., & Rodriguez-Lara, I. (2018). Cheating, incentives, and money manipulation. *Experimental Economics*. Advance online publication. <http://dx.doi.org/10.1007/s10683-018-9584-1>

- Cherry, T. L., Frykblom, P., & Shogren, J. F. (2002). Hardnose the dictator. *The American Economic Review*, 92, 1218–1221. <http://dx.doi.org/10.1257/00028280260344740>
- Cheung, M. W. L. (2015). metaSEM: An R package for meta-analysis using structural equation modeling. *Frontiers in Psychology*, 5, 1521. <http://dx.doi.org/10.3389/fpsyg.2014.01521>
- Childs, J. (2012a). Demonstrating the need for effective business ethics. An alternative approach. *Business and Society Review*, 117, 221–232. <http://dx.doi.org/10.1111/j.1467-8594.2012.00406.x>
- Childs, J. (2012b). Gender differences in lying. *Economics Letters*, 114, 147–149. <http://dx.doi.org/10.1016/j.econlet.2011.10.006>
- Childs, J. (2013). Personal characteristics and lying. An experimental investigation. *Economics Letters*, 121, 425–427. <http://dx.doi.org/10.1016/j.econlet.2013.09.005>
- Chou, E. Y. (2015). What's in a name? The toll e-signatures take on individual honesty. *Journal of Experimental Social Psychology*, 61, 84–95. <http://dx.doi.org/10.1016/j.jesp.2015.07.010>
- Cialdini, R. B., Reno, R. R., & Kallgren, C. A. (1990). A focus theory of normative conduct: Recycling the concept of norms to reduce littering in public places. *Journal of Personality and Social Psychology*, 58, 1015–1026. <http://dx.doi.org/10.1037/0022-3514.58.6.1015>
- Clot, S., Grolleau, G., & Ibanez, L. (2014). Smug alert! Exploring self-licensing behavior in a cheating game. *Economics Letters*, 123, 191–194. <http://dx.doi.org/10.1016/j.econlet.2014.01.039>
- Cohen, T. R., Gunia, B. C., Kim-Jun, S. Y., & Murnighan, J. K. (2009). Do groups lie more than individuals? Honesty and deception as a function of strategic self-interest. *Journal of Experimental Social Psychology*, 45, 1321–1324. <http://dx.doi.org/10.1016/j.jesp.2009.08.007>
- Cohen, T. R., Wolf, S. T., Panter, A. T., & Insko, C. A. (2011). Introducing the GASP scale: A new measure of guilt and shame proneness. *Journal of Personality and Social Psychology*, 100, 947–966. <http://dx.doi.org/10.1037/a0022641>
- Cohn, A., Maréchal, M. A., & Noll, T. (2015). Bad boys: How criminal identity salience affects rule violation. *The Review of Economic Studies*, 82, 1289–1308. <http://dx.doi.org/10.1093/restud/rdrv025>
- Conrads, J., Ellenberger, M., Irlenbusch, B., Ohms, E. N., Rilke, R. M., & Walkowitz, G. (2016). Team goal incentives and individual lying behavior. *Die Betriebswirtschaft*, 76, 103–123.
- Conrads, J., Irlenbusch, B., Rilke, R. M., Schielke, A., & Walkowitz, G. (2014). Honesty in tournaments. *Economics Letters*, 123, 90–93. <http://dx.doi.org/10.1016/j.econlet.2014.01.026>
- Conrads, J., Irlenbusch, B., Rilke, R. M., & Walkowitz, G. (2013). Lying and team incentives. *Journal of Economic Psychology*, 34, 1–7. <http://dx.doi.org/10.1016/j.joep.2012.10.011>
- Conrads, J., & Lotz, S. (2015). The effect of communication channels on dishonest behavior. *Journal of Behavioral and Experimental Economics*, 58, 88–93. <http://dx.doi.org/10.1016/j.socec.2015.06.006>
- Crone, D. L., & Levy, N. L. (2018). Are free will believers nicer people? (Four studies suggest not). *Social Psychological & Personality Science*. Advance online publication. Retrieved from <https://psyarxiv.com/zpj5x/>
- Cropanzano, R., & Mitchell, M. S. (2005). Social exchange theory: An Interdisciplinary review. *Journal of Management*, 31, 874–900. <http://dx.doi.org/10.1177/0149206305279602>
- Crosen, R., & Gneezy, U. (2009). Gender differences in preferences. *Journal of Economic Literature*, 47, 448–474. <http://dx.doi.org/10.1257/jel.47.2.448>
- Dai, Z., Galeotti, F., & Villeval, M. C. (2018). Cheating in the lab predicts fraud in the field. An experiment in public transportations. *Management Science*, 64, 1081–1100. <http://dx.doi.org/10.1287/mnsc.2016.2616>
- Dana, J., Weber, R. A., & Kuang, J. X. (2007). Exploiting moral wiggle room. Experiments demonstrating an illusory preference for fairness. *Economic Theory*, 33, 67–80. <http://dx.doi.org/10.1007/s00199-006-0153-z>
- Darley, J. M., & Latané, B. (1968). Bystander intervention in emergencies: Diffusion of responsibility. *Journal of Personality and Social Psychology*, 8, 377–383. <http://dx.doi.org/10.1037/h0025589>
- Denson, T. F., DeWall, C. N., & Finkel, E. J. (2012). Self-control and aggression. *Current Directions in Psychological Science*, 21, 20–25. <http://dx.doi.org/10.1177/0963721411429451>
- Dhimi, M. K., Hertwig, R., & Hoffrage, U. (2004). The role of representative design in an ecological approach to cognition. *Psychological Bulletin*, 130, 959–988. <http://dx.doi.org/10.1037/0033-2909.130.6.959>
- Dieckmann, A., Grimm, V., Unfried, M., Utikal, V., & Valmasoni, L. (2016). On trust in honesty and volunteering among Europeans. Cross-country evidence on perceptions and behavior. *European Economic Review*, 90, 225–253. <http://dx.doi.org/10.1016/j.euroecorev.2016.01.011>
- Dreber, A., & Johannesson, M. (2008). Gender differences in deception. *Economics Letters*, 99, 197–199. <http://dx.doi.org/10.1016/j.econlet.2007.06.027>
- Dubois, D., Rucker, D. D., & Galinsky, A. D. (2015). Social class, power, and selfishness: When and why upper and lower class individuals behave unethically. *Journal of Personality and Social Psychology*, 108, 436–449. <http://dx.doi.org/10.1037/pspi0000008>
- Duncan, D., & Li, D. (2018). Liar liar: Experimental evidence of the effect of confirmation-reports on dishonesty. *Southern Economic Journal*, 84, 742–770. <http://dx.doi.org/10.1002/soej.12244>
- Duval, S., & Tweedie, R. (2000). A non-parametric “trim and fill” method of assessing publication bias in meta-analysis. *Journal of the American Statistical Association*, 95, 89–98.
- Egger, M., Davey Smith, G., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *British Medical Journal (Clinical Research Ed.)*, 315, 629–634. <http://dx.doi.org/10.1136/bmj.315.7109.629>
- Erat, S. (2013). Avoiding lying. The case of delegated deception. *Journal of Economic Behavior & Organization*, 93, 273–278. <http://dx.doi.org/10.1016/j.jebo.2013.03.035>
- Erat, S., & Gneezy, U. (2012). White lies. *Management Science*, 58, 723–733. <http://dx.doi.org/10.1287/mnsc.1110.1449>
- Eriksson, K., Strimling, P., & Coultas, J. C. (2015). Bidirectional associations between descriptive and injunctive norms. *Organizational Behavior and Human Decision Processes*, 129, 59–69. <http://dx.doi.org/10.1016/j.obhdp.2014.09.011>
- Ex-PM Olmert released from prison after 16 months. (2017, July 2). *The Times of Israel*. Retrieved from <http://www.timesofisrael.com/ex-pm-olmert-to-go-free-after-16-months-in-prison/>
- Ezquerro, L., Kolev, G. I., & Rodriguez-Lara, I. (2018). Gender differences in cheating: Loss vs. gain framing. *Economics Letters*, 163, 46–49. <http://dx.doi.org/10.1016/j.econlet.2017.11.016>
- Faravelli, M., Friesen, L., & Gangadharan, L. (2015). Selection, tournaments, and dishonesty. *Journal of Economic Behavior & Organization*, 110, 160–175. <http://dx.doi.org/10.1016/j.jebo.2014.10.019>
- Fischbacher, U., & Föllmi-Heusi, F. (2013). Lies in disguise. An experimental study on cheating. *Journal of the European Economic Association*, 11, 525–547. <http://dx.doi.org/10.1111/jeea.12014>
- Fischbacher, U., & Heusi, F. (2008). Lies in disguise. An experimental study on cheating (TWI Research Paper). Retrieved from <http://econpapers.repec.org/paper/twirespas/0040.htm>
- Fosgaard, T. R., Bucciol, A., Hansen, L. G., & Piovesan, M. (2018). Students cheat more. Comparing dishonesty of a student and a representative sample in the laboratory. *The Scandinavian Journal of Economics*. Advance online publication. <http://dx.doi.org/10.1111/sjoe.12326>
- Fosgaard, T. R., Hansen, L. G., & Piovesan, M. (2013). Separating will from grace. An experiment on conformity and awareness in cheating. *Journal of Economic Behavior & Organization*, 93, 279–284. <http://dx.doi.org/10.1016/j.jebo.2013.03.027>

- Frey, B. S., Pommerehne, W. W., & Gygi, B. (1993). Economic indoctrination or selection? Some empirical results. *The Journal of Economic Education*, 24, 271–281. <http://dx.doi.org/10.1080/00220485.1993.10844799>
- Friesen, L., & Gangadharan, L. (2012). Individual level evidence of dishonesty and the gender effect. *Economics Letters*, 117, 624–626. <http://dx.doi.org/10.1016/j.econlet.2012.08.005>
- Friesen, L., & Gangadharan, L. (2013). Designing self-reporting regimes to encourage truth telling: An experimental study. *Journal of Economic Behavior & Organization*, 94, 90–102. <http://dx.doi.org/10.1016/j.jebo.2013.08.007>
- Gächter, S., & Schulz, J. F. (2016). Intrinsic honesty and the prevalence of rule violations across societies. *Nature*, 531, 496–499. <http://dx.doi.org/10.1038/nature17160>
- Gamliel, E., & Peer, E. (2013). Explicit risk of getting caught does not affect unethical behavior. *Journal of Applied Social Psychology*, 43, 1281–1288. <http://dx.doi.org/10.1111/jasp.12091>
- Gawn, G., & Innes, R. (2018). Do lies erode trust? *International Economic Review*, 59, 137–161. <http://dx.doi.org/10.1111/iere.12265>
- Gerlach, P. (2017). The games economists play: Why economics students behave more selfishly than other students. *PLoS ONE*, 12, e0183814. <http://dx.doi.org/10.1371/journal.pone.0183814>
- Gino, F. (2015). Understanding ordinary unethical behavior. Why people who value morality act immorally. *Current Opinion in Behavioral Sciences*, 3, 107–111. <http://dx.doi.org/10.1016/j.cobeha.2015.03.001>
- Gino, F., & Ariely, D. (2012). The dark side of creativity: Original thinkers can be more dishonest. *Journal of Personality and Social Psychology*, 102, 445–459. <http://dx.doi.org/10.1037/a0026406>
- Gino, F., & Ariely, D. (2016). Dishonesty explained. What leads moral people act immorally. In A. G. Miller (Ed.), *The social psychology of good and evil* (pp. 322–344). New York, NY: Guilford Press.
- Gino, F., Ayal, S., & Ariely, D. (2009). Contagion and differentiation in unethical behavior: The effect of one bad apple on the barrel. *Psychological Science*, 20, 393–398. <http://dx.doi.org/10.1111/j.1467-9280.2009.02306.x>
- Gino, F., Ayal, S., & Ariely, D. (2013). Self-serving altruism? The lure of unethical actions that benefit others. *Journal of Economic Behavior & Organization*, 93, 285–292. <http://dx.doi.org/10.1016/j.jebo.2013.04.005>
- Gino, F., & Galinsky, A. D. (2012). Vicarious dishonesty. When psychological closeness creates distance from one's moral compass. *Organizational Behavior and Human Decision Processes*, 119, 15–26. <http://dx.doi.org/10.1016/j.obhdp.2012.03.011>
- Gino, F., Krupka, E. L., & Weber, R. A. (2013). License to cheat. Voluntary regulation and ethical behavior. *Management Science*, 59, 2187–2203. <http://dx.doi.org/10.1287/mnsc.1120.1699>
- Gino, F., & Margolis, J. D. (2011). Bringing ethics into focus. How regulatory focus and risk preferences influence (un)ethical behavior. *Organizational Behavior and Human Decision Processes*, 115, 145–156. <http://dx.doi.org/10.1016/j.obhdp.2011.01.006>
- Gino, F., & Mogilner, C. (2014). Time, money, and morality. *Psychological Science*, 25, 414–421. <http://dx.doi.org/10.1177/0956797613506438>
- Gino, F., Norton, M. I., & Ariely, D. (2010). The counterfeit self: The deceptive costs of faking it. *Psychological Science*, 21, 712–720. <http://dx.doi.org/10.1177/0956797610366545>
- Gino, F., Schweitzer, M. E., Mead, N. L., & Ariely, D. (2011). Unable to resist temptation. How self-control depletion promotes unethical behavior. *Organizational Behavior and Human Decision Processes*, 115, 191–203. <http://dx.doi.org/10.1016/j.obhdp.2011.03.001>
- Gino, F., & Wiltermuth, S. S. (2014). Evil genius? How dishonesty can lead to greater creativity. *Psychological Science*, 25, 973–981. <http://dx.doi.org/10.1177/0956797614520714>
- Glätzle-Rützler, D., & Lерgetporer, P. (2015). Lying and age. An experimental study. *Journal of Economic Psychology*, 46, 12–25. <http://dx.doi.org/10.1016/j.joep.2014.11.002>
- Gneezy, U. (2005). Deception: The role of consequences. *The American Economic Review*, 95, 384–394. <http://dx.doi.org/10.1257/0002828053828662>
- Gneezy, U., Kajackaite, A., & Sobel, J. (2018). Lying aversion and the size of the lie. *The American Economic Review*, 108, 419–453. <http://dx.doi.org/10.1257/aer.20161553>
- Gravert, C. (2013). How luck and performance affect stealing. *Journal of Economic Behavior & Organization*, 93, 301–304. <http://dx.doi.org/10.1016/j.jebo.2013.03.026>
- Greene, J. D., & Paxton, J. M. (2009). Patterns of neural activity associated with honest and dishonest moral decisions. *Proceedings of the National Academy of Sciences of the United States of America*, 106, 12506–12511. <http://dx.doi.org/10.1073/pnas.0900152106>
- Grolleau, G., Kocher, M. G., & Sutan, A. (2016). Cheating and loss aversion. Do people cheat more to avoid a loss? *Management Science*, 62, 3428–3438. <http://dx.doi.org/10.1287/mnsc.2015.2313>
- Gu, J., Zhong, C.-B., & Page-Gould, E. (2013). Listen to your heart: When false somatic feedback shapes moral behavior. *Journal of Experimental Psychology: General*, 142, 307–312. <http://dx.doi.org/10.1037/a0029549>
- Gunia, B. C., Barnes, C. M., & Sah, S. (2014). The morality of larks and owls: Unethical behavior depends on chronotype as well as time of day. *Psychological Science*, 25, 2272–2274. <http://dx.doi.org/10.1177/0956797614541989>
- Gunia, B. C., Wang, L., Huang, L., Wang, J., & Murnighan, J. K. (2012). Contemplation and conversation. Subtle influences on moral decision making. *Academy of Management Journal*, 55, 13–33. <http://dx.doi.org/10.5465/amj.2009.0873>
- Gylfason, H. F., Arnardottir, A. A., & Kristinsson, K. (2013). More on gender differences in lying. *Economics Letters*, 119, 94–96. <http://dx.doi.org/10.1016/j.econlet.2013.01.027>
- Gylfason, H. F., Halldorsson, F., & Kristinsson, K. (2016). Personality in Gneezy's cheap talk game. The interaction between honesty-humility and extraversion in predicting deceptive behavior. *Personality and Individual Differences*, 96, 222–226. <http://dx.doi.org/10.1016/j.paid.2016.02.075>
- Hendershott, A., Drinan, P., & Cross, M. (1999). Gender and academic integrity. *Journal of College Student Development*, 40, 345–354.
- Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world? *Behavioral and Brain Sciences*, 33, 61–83. <http://dx.doi.org/10.1017/S0140525X0999152X>
- Hershfield, H. E., Cohen, T. R., & Thompson, L. (2012). Short horizons and tempting situations: Lack of continuity to our future selves leads to unethical decision making and behavior. *Organizational Behavior and Human Decision Processes*, 117, 298–310. <http://dx.doi.org/10.1016/j.obhdp.2011.11.002>
- Hertwig, R., & Ortmann, A. (2001). Experimental practices in economics: A methodological challenge for psychologists? *Behavioral and Brain Sciences*, 24, 383–451.
- Hertwig, R., & Ortmann, A. (2008a). Deception in experiments: Revisiting the arguments in its defense. *Ethics & Behavior*, 18, 59–92. <http://dx.doi.org/10.1080/10508420701712990>
- Hertwig, R., & Ortmann, A. (2008b). Deception in social psychological experiments: Two misconceptions and a research agenda. *Social Psychology Quarterly*, 71, 222–227. <http://dx.doi.org/10.1177/019027250807100304>
- Hilbig, B. E., & Hessler, C. M. (2013). What lies beneath. How the distance between truth and lie drives dishonesty. *Journal of Experimental Social Psychology*, 49, 263–266. <http://dx.doi.org/10.1016/j.jesp.2012.11.010>

- Hilbig, B. E., Moshagen, M., & Zettler, I. (2016). Prediction consistency: A test of the equivalence assumption across different indicators of the same construct. *European Journal of Personality, 30*, 637–647. <http://dx.doi.org/10.1002/per.2085>
- Hilbig, B. E., & Thielmann, I. (2017). Does everyone have a price? On the role of payoff magnitude for ethical decision making. *Cognition, 163*, 15–25. <http://dx.doi.org/10.1016/j.cognition.2017.02.011>
- Hilbig, B. E., & Zettler, I. (2015). When the cat's away, some mice will play. A basic trait account of dishonest behavior. *Journal of Research in Personality, 57*, 72–88. <http://dx.doi.org/10.1016/j.jrp.2015.04.003>
- Hilbig, B. E., & Zettler, I. (2018). *Unpublished working paper*.
- Hildreth, J. A. D., Gino, F., & Bazerman, M. (2016). Blind loyalty? When group loyalty makes us see evil or engage in it. *Organizational Behavior and Human Decision Processes, 132*, 16–36. <http://dx.doi.org/10.1016/j.obhdp.2015.10.001>
- Hoffmann, A., Diederhofen, B., Verschuere, B., & Musch, J. (2015). A strong validation of the crosswise model using experimentally-induced cheating behavior. *Experimental Psychology, 62*, 403–414. <http://dx.doi.org/10.1027/1618-3169/a000304>
- Holm, H. J., & Kawagoe, T. (2010). Face-to-face lying. An experimental study in Sweden and Japan. *Journal of Economic Psychology, 31*, 310–321. <http://dx.doi.org/10.1016/j.joep.2010.01.001>
- Houser, D., List, J. A., Piovesan, M., Samek, A., & Winter, J. (2016). Dishonesty: From parents to children. *European Economic Review, 82*, 242–254. <http://dx.doi.org/10.1016/j.euroecorev.2015.11.003>
- Houser, D., Vetter, S., & Winter, J. (2012). Fairness and cheating. *European Economic Review, 56*, 1645–1655. <http://dx.doi.org/10.1016/j.euroecorev.2012.08.001>
- Huedo-Medina, T. B., Sánchez-Meca, J., Marín-Martínez, F., & Botella, J. (2006). Assessing heterogeneity in meta-analysis: Q statistic or I2 index? *Psychological Methods, 11*, 193–206. <http://dx.doi.org/10.1037/1082-989X.11.2.193>
- Hugh-Jones, D. (2016). Honesty, beliefs about honesty, and economic growth in 15 countries. *Journal of Economic Behavior & Organization, 127*, 99–114. <http://dx.doi.org/10.1016/j.jebo.2016.04.012>
- Hurkens, S., & Kartik, N. (2009). Would I lie to you? On social preferences and lying aversion. *Experimental Economics, 12*, 180–192. <http://dx.doi.org/10.1007/s10683-008-9208-2>
- Innes, R., & Mitra, I. (2013). Is dishonesty contagious? *Economic Inquiry, 51*, 722–734. <http://dx.doi.org/10.1111/j.1465-7295.2012.00470.x>
- International Monetary Fund. (2015). World economic and financial surveys: World Economic Outlook Database. Retrieved from <https://www.imf.org/external/pubs/ft/weo/2015/01/weodata/index.aspx>
- Jacobsen, C., Fosgaard, T. R., & Pascual-Ezama, D. (2018). Why do we lie? A practical guide to the dishonesty literature. *Journal of Economic Surveys, 32*, 357–387. <http://dx.doi.org/10.1111/joes.12204>
- Jacobsen, C., & Piovesan, M. (2016). Tax me if you can. An artificial field experiment on dishonesty. *Journal of Economic Behavior & Organization, 124*, 7–14. <http://dx.doi.org/10.1016/j.jebo.2015.09.009>
- Josef, A. K., Richter, D., Samanez-Larkin, G. R., Wagner, G. G., Hertwig, R., & Mata, R. (2016). Stability and change in risk-taking propensity across the adult life span. *Journal of Personality and Social Psychology, 111*, 430–450. <http://dx.doi.org/10.1037/pspp0000090>
- Kajackaite, A. (2018). Lying about luck versus lying about performance. *Journal of Economic Behavior, 153*, 194–199. <http://dx.doi.org/10.1016/j.jebo.2018.07.010>
- Kajackaite, A., & Gneezy, U. (2017). Incentives and cheating. *Games and Economic Behavior, 102*, 433–444. <http://dx.doi.org/10.1016/j.geb.2017.01.015>
- Kallgren, C. A., Reno, R. R., & Cialdini, R. B. (2000). A focus theory of normative conduct: When norms do and do not affect behavior. *Personality and Social Psychology Bulletin, 26*, 1002–1012. <http://dx.doi.org/10.1177/01461672002610009>
- Keizer, K., Lindenberg, S., & Steg, L. (2008). The spreading of disorder. *Science, 322*, 1681–1685. <http://dx.doi.org/10.1126/science.1161405>
- Kerschbamer, R., Neururer, D., & Gruber, A. (2017). Do the altruists lie less? Retrieved from https://www.uibk.ac.at/economics/bbl/cv_papiere/101-do-the-altruists-lie-less-wp-2017-18.pdf
- Kilduff, G., Galinsky, A., Gallo, E., & Reade, J. (2016). Whatever it takes to win: Rivalry increases unethical behavior. *Academy of Management Journal, 59*, 1508–1534. <http://dx.doi.org/10.5465/amj.2014.0545>
- Kimmel, A. J. (1996). *Ethical issues in behavioral research: A survey*. Cambridge, MA: Blackwell.
- Kimmel, A. J. (2007). *Ethical issues in behavioral research: Basic and applied perspectives, 2nd ed.* Malden, MA: Blackwell.
- Kirchgässner, G. (2005). (Why) are economists different? *European Journal of Political Economy, 21*, 543–562. <http://dx.doi.org/10.1016/j.ejpoleco.2005.05.003>
- Klein, S. A., Thielmann, I., Hilbig, B. E., & Zettler, I. (2018). *Unpublished working paper*.
- Kleinlogel, E. P., Dietz, J., & Antonakis, J. (2018). Lucky, competent, or just a cheat? Interactive effects of honesty-humility and moral cues on cheating behavior. *Personality and Social Psychology Bulletin, 44*, 158–172. <http://dx.doi.org/10.1177/0146167217733071>
- Klimm, F. (2018, January). *Suspicious success-Cheating, inequality acceptance and political preferences* (Munich Discussion Paper No. 2018–1). Retrieved from <http://epub.uni-muenchen.de/41998/>
- Korbel, V. (2017). Do we lie in groups? An experimental evidence. *Applied Economics Letters, 24*, 1107–1111. <http://dx.doi.org/10.1080/13504851.2016.1259734>
- Kouchaki, M., Gino, F., & Jami, A. (2014). The burden of guilt: Heavy backpacks, light snacks, and enhanced morality. *Journal of Experimental Psychology: General, 143*, 414–424. <http://dx.doi.org/10.1037/a0031769>
- Kouchaki, M., & Smith, I. H. (2014). The morning morality effect: The influence of time of day on unethical behavior. *Psychological Science, 25*, 95–102. <http://dx.doi.org/10.1177/0956797613498099>
- Kouchaki, M., Smith-Crowe, K., Brief, A. P., & Sousa, C. (2013). Seeing green. Mere exposure to money triggers a business decision frame and unethical outcomes. *Organizational Behavior and Human Decision Processes, 121*, 53–61. <http://dx.doi.org/10.1016/j.obhdp.2012.12.002>
- Kouchaki, M., & Wareham, J. (2015). Excluded and behaving unethically: Social exclusion, physiological responses, and unethical behavior. *Journal of Applied Psychology, 100*, 547–556. <http://dx.doi.org/10.1037/a0038034>
- Lammers, J., Stapel, D. A., & Galinsky, A. D. (2010). Power increases hypocrisy: Moralizing in reasoning, immorality in behavior. *Psychological Science, 21*, 737–744. <http://dx.doi.org/10.1177/0956797610368810>
- Lee, J. J., Gino, F., Jin, E. S., Rice, L. K., & Josephs, R. A. (2015). Hormones and ethics: Understanding the biological basis of unethical conduct. *Journal of Experimental Psychology: General, 144*, 891–897. <http://dx.doi.org/10.1037/xge0000099>
- Lee, J. J., Im, D. K., Parmar, B. L., & Gino, F. (2015, February). Thick as thieves? Dishonest behavior and egocentric social networks (Harvard Business School NOM Unit Working Paper). <http://dx.doi.org/10.2139/ssrn.2563196>
- Leibbrandt, A., Maitra, P., & Neelim, A. (2017, March). *Large stakes and little honesty? Experimental evidence from a developing country* (Discussion Paper 13/17). Retrieved from <https://www.monash.edu/business/economics/research/publications/publications2/1317honestyleibbrandtmaitranceelim.pdf>
- Lerner, J. S., & Keltner, D. (2001). Fear, anger, and risk. *Journal of Personality and Social Psychology, 81*, 146–159. <http://dx.doi.org/10.1037/0022-3514.81.1.146>
- Lewis, A., Bardis, A., Flint, C., Mason, C., Smith, N., Tickle, C., & Zinser, J. (2012). Drawing the line somewhere. An experimental study of moral

- compromise. *Journal of Economic Psychology*, 33, 718–725. <http://dx.doi.org/10.1016/j.joep.2012.01.005>
- “Luiz Inácio Lula da Silva. The Rise and fall of a Brazilian leader”. (2017, July 12). *The New York Times*. Retrieved from <https://www.nytimes.com/2017/07/12/world/americas/luiz-inacio-lula-da-silva-the-rise-and-fall-of-a-brazilian-leader.html>
- “Nicolas Sarkozy to face trial for corruption and influence peddling”. (2018, March 29). *The Guardian*.
- Lohse, T., Simon, S. A., & Konrad, K. A. (2018). Deception under time pressure: Conscious decision or a problem of awareness? *Journal of Economic Behavior & Organization*, 146, 31–42. <http://dx.doi.org/10.1016/j.jebo.2017.11.026>
- López-Pérez, R., & Spiegelman, E. (2012, January). *Do economists lie more?* (Working paper No. 1885–6888). Madrid, Spain. Retrieved from <http://www.uam.es/departamentos/economicas/analecon/especifica/mimeo/wp20124.pdf>
- López-Pérez, R., & Spiegelman, E. (2013). Why do people tell the truth? Experimental evidence for pure lie aversion. *Experimental Economics*, 16, 233–247. <http://dx.doi.org/10.1007/s10683-012-9324-x>
- Lundquist, T., Ellingsen, T., Gribbe, E., & Johannesson, M. (2009). The aversion to lying. *Journal of Economic Behavior & Organization*, 70, 81–92. <http://dx.doi.org/10.1016/j.jebo.2009.02.010>
- Mai, K. M., Ellis, A. P. J., & Welsh, D. T. (2015). The gray side of creativity. Exploring the role of activation in the link between creative personality and unethical behavior. *Journal of Experimental Social Psychology*, 60, 76–85. <http://dx.doi.org/10.1016/j.jesp.2015.05.004>
- Mata, R., Josef, A. K., & Hertwig, R. (2016). Propensity for risk taking across the life span and around the globe. *Psychological Science*, 27, 231–243. <http://dx.doi.org/10.1177/0956797615617811>
- Mazar, N., Amir, O., & Ariely, D. (2008). The dishonesty of honest people. A theory of self-concept maintenance. *Journal of Marketing Research*, 45, 633–644. <http://dx.doi.org/10.1509/jmkr.45.6.633>
- McShane, B. B., Böckenholt, U., & Hansen, K. T. (2016). Adjusting for publication bias in meta-analysis: An evaluation of selection methods and some cautionary notes. *Perspectives on Psychological Science*, 11, 730–749. <http://dx.doi.org/10.1177/1745691616662243>
- Mead, N. L., Baumeister, R. F., Gino, F., Schweitzer, M. E., & Ariely, D. (2009). Too tired to tell the truth: Self-control resource depletion and dishonesty. *Journal of Experimental Social Psychology*, 45, 594–597. <http://dx.doi.org/10.1016/j.jesp.2009.02.004>
- Milgram, S. (1974). *Obedience to authority. An experimental view*. New York, NY: Harper Collins.
- Moshagen, M., Hilbig, B. E., & Zettler, I. (2018). The dark core of personality. *Psychological Review*. Advance online publication. <http://dx.doi.org/10.1037/rev0000111>
- Motro, D., Ordóñez, L. D., Pittarello, A., & Welsh, D. T. (2016). Investigating the effects of anger and guilt on unethical behavior: A dual-process approach. *Journal of Business Ethics*, 152, 133–148.
- Muehlheusser, G., Roider, A., & Wallmeier, N. (2015). Gender differences in honesty. Groups versus individuals. *Economics Letters*, 128, 25–29. <http://dx.doi.org/10.1016/j.econlet.2014.12.019>
- Muñoz-Izquierdo, N., de Liaño, B. G.-G., Rin-Sánchez, F. D., & Pascual-Ezama, D. (2014, December). *Economists: Cheaters with altruistic instincts* (Munich Personal RePEc Archive). Retrieved from <https://mpra.ub.uni-muenchen.de/60678/>
- Ortmann, A., & Hertwig, R. (2002). The costs of deception: Evidence from psychology. *Experimental Economics*, 5, 111–131. <http://dx.doi.org/10.1023/A:1020365204768>
- Pascual-Ezama, D., Fosgaard, T. R., Cardenas, J. C., Kujal, P., Veszteg, R., Gil-Gómez de Liaño, B., . . . Brañas-Garza, P. (2015). Context-dependent cheating. Experimental evidence from 16 countries. *Journal of Economic Behavior & Organization*, 116, 379–386. <http://dx.doi.org/10.1016/j.jebo.2015.04.020>
- Peer, E., Acquisti, A., & Shalvi, S. (2014). “I cheated, but only a little”: Partial confessions to unethical behavior. *Journal of Personality and Social Psychology*, 106, 202–217. <http://dx.doi.org/10.1037/a0035392>
- Peer, E., Brandimarte, L., Samat, S., & Acquisti, A. (2017). Beyond the Turk: Alternative platforms for crowdsourcing behavioral research. *Journal of Experimental Social Psychology*, 70, 153–163. <http://dx.doi.org/10.1016/j.jesp.2017.01.006>
- Peeters, R., Vorsatz, M., & Walzl, M. (2015). Beliefs and truth-telling: A laboratory experiment. *Journal of Economic Behavior & Organization*, 113, 1–12. <http://dx.doi.org/10.1016/j.jebo.2015.02.009>
- Piff, P. K., Stancato, D. M., Côté, S., Mendoza-Denton, R., & Keltner, D. (2012). Higher social class predicts increased unethical behavior. *Proceedings of the National Academy of Sciences of the United States of America*, 109, 4086–4091. <http://dx.doi.org/10.1073/pnas.1118373109>
- Ploner, M., & Regner, T. (2013). Self-image and moral balancing. An experimental analysis. *Journal of Economic Behavior & Organization*, 93, 374–383. <http://dx.doi.org/10.1016/j.jebo.2013.03.030>
- Rasmußen, A. (2015). Reporting behavior. A literature review of experimental studies. *Central European Journal of Operations Research*, 23, 283–311. <http://dx.doi.org/10.1007/s10100-014-0379-y>
- R Development Core Team. (2008). R: A language and environment for statistical computing [Computer Software]. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from <http://www.r-project.org>
- Rhine, W. J. (2008, June). *Culture, self-orientation, and reward structure effects. Measuring cheating behavior in China and the USA*. (SSRN Working paper series). Retrieved from https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1931751
- Rigdon, M. L., & D’Esterre, A. (2014). The effects of competition on the nature of cheating behavior. *Southern Economic Journal*, 81, 1012–1024.
- Rigdon, M. L., & D’Esterre, A. (2017). Sabotaging another: Priming competition increases cheating behavior in tournaments. *Southern Economic Journal*, 84, 456–473. <http://dx.doi.org/10.1002/soej.12232>
- Roberts, B. W., Walton, K. E., & Viechtbauer, W. (2006). Patterns of mean-level change in personality traits across the life course: A meta-analysis of longitudinal studies. *Psychological Bulletin*, 132, 1–25. <http://dx.doi.org/10.1037/0033-2909.132.1.1>
- Rode, J. (2010). Truth and trust in communication: Experiments on the effect of a competitive context. *Games and Economic Behavior*, 68, 325–338. <http://dx.doi.org/10.1016/j.geb.2009.05.008>
- Roeser, K., McGregor, V. E., Stegmaier, S., Mathew, J., Kübler, A., & Meule, A. (2016). The dark triad of personality and unethical behavior at different times of day. *Personality and Individual Differences*, 88, 73–77. <http://dx.doi.org/10.1016/j.paid.2015.09.002>
- Rosenbaum, S. M., Billinger, S., & Stieglitz, N. (2014). Let’s be honest: A review of experimental evidence of honesty and truth-telling. *Journal of Economic Psychology*, 45, 181–196. <http://dx.doi.org/10.1016/j.joep.2014.10.002>
- Ruffle, B. J., & Tobol, Y. (2014). Honest on Mondays: Honesty and the temporal separation between decisions and payoffs. *European Economic Review*, 65, 126–135. <http://dx.doi.org/10.1016/j.eurocorev.2013.11.004>
- Ruffle, B. J., & Tobol, Y. (2017). Clever enough to tell the truth. *Experimental Economics*, 20, 130–155. <http://dx.doi.org/10.1007/s10683-016-9479-y>
- Schindler, S., & Pfattheicher, S. (2017). The frame of the game: Loss-framing increases dishonest behavior. *Journal of Experimental Social Psychology*, 69, 172–177. <http://dx.doi.org/10.1016/j.jesp.2016.09.009>
- Schurr, A., & Ritov, I. (2016). Winning a competition predicts dishonest behavior. *Proceedings of the National Academy of Sciences of the United States of America*, 113, 1754–1759. <http://dx.doi.org/10.1073/pnas.1515102113>
- Schwarzer, G. (2007). meta: An R package for meta-analysis. *R News*, 7, 40–45.

- Schweitzer, M. E., & Hsee, C. K. (2002). Stretching the truth: Elastic justification and motivated communication of uncertain information. *Journal of Risk and Uncertainty*, 25, 185–201. <http://dx.doi.org/10.1023/A:1020647814263>
- Scott, J. H., Jr., & Rothman, M. P. (1975). The effect of an introductory economics course on student political attitudes. *The Journal of Economic Education*, 6, 107–112. <http://dx.doi.org/10.1080/00220485.1975.10845410>
- Shalvi, S., Dana, J., Handgraaf, M. J. J., & De Dreu, C. K. W. (2011). Justified ethicality: Observing desired counterfactuals modifies ethical perceptions and behavior. *Organizational Behavior and Human Decision Processes*, 115, 181–190. <http://dx.doi.org/10.1016/j.obhdp.2011.02.001>
- Shalvi, S., Eldar, O., & Bereby-Meyer, Y. (2012). Honesty requires time (and lack of justifications). *Psychological Science*, 23, 1264–1270. <http://dx.doi.org/10.1177/0956797612443835>
- Shalvi, S., Handgraaf, M. J. J., & De Dreu, C. K. W. (2011). Ethical manoeuvring: Why people avoid both major and minor lies. *British Journal of Management*, 22, S16–S27. <http://dx.doi.org/10.1111/j.1467-8551.2010.00709.x>
- Shalvi, S., & Leiser, D. (2013). Moral firmness. *Journal of Economic Behavior & Organization*, 93, 400–407. <http://dx.doi.org/10.1016/j.jebo.2013.03.014>
- Shu, L. L., & Gino, F. (2012). Sweeping dishonesty under the rug: How unethical actions lead to forgetting of moral rules. *Journal of Personality and Social Psychology*, 102, 1164–1177. <http://dx.doi.org/10.1037/a0028381>
- Shu, L. L., Gino, F., & Bazerman, M. H. (2011). Dishonest deed, clear conscience: When cheating leads to moral disengagement and motivated forgetting. *Personality and Social Psychology Bulletin*, 37, 330–349. <http://dx.doi.org/10.1177/0146167211398138>
- Shu, L. L., Mazar, N., Gino, F., Ariely, D., & Bazerman, M. H. (2012). Signing at the beginning makes ethics salient and decreases dishonest self-reports in comparison to signing at the end. *Proceedings of the National Academy of Sciences of the United States of America*, 109, 15197–15200. <http://dx.doi.org/10.1073/pnas.1209746109>
- Silverman, I., Shulman, A. D., & Wiesenfeld, D. L. (1970). Effects of deceiving and debriefing psychological subjects on performance in later experiments. *Journal of Personality and Social Psychology*, 14, 203–212. <http://dx.doi.org/10.1037/h0028852>
- Sutter, M. (2009). Deception through telling the truth?! Experimental evidence from individuals and teams. *Economic Journal*, 119, 47–60. <http://dx.doi.org/10.1111/j.1468-0297.2008.02205.x>
- Tenbrunsel, A. E., & Messick, D. M. (2004). Ethical fading: The role of self-deception in unethical behavior. *Social Justice Research*, 17, 223–236. <http://dx.doi.org/10.1023/B:SORE.0000027411.35832.53>
- Ter Meer, J. (2014, July). *The indirect effect of monetary incentives on deception* (CGS Working Paper). Retrieved from http://www.cgs.uni-koeln.de/fileadmin/wiso_fak/cgs/pdf/working_paper/cgswp_05-04.pdf
- Thielmann, I., & Hilbig, B. E. (2018a). *No gain without pain: The psychological costs of dishonesty*. Advance online publication. <http://dx.doi.org/10.1016/j.joep.2018.06.001>
- Thielmann, I., & Hilbig, B. E. (2018b). *Cheating, personality, and incentives*. Manuscript in preparation.
- Thielmann, I., Hilbig, B. E., Zettler, I., & Moshagen, M. (2017). On measuring the sixth basic personality dimension: A comparison between HEXACO honesty-humility and Big Six honesty-propriety. *Assessment*, 24, 1024–1036. <http://dx.doi.org/10.1177/1073191116638411>
- Transparency International. (2017). *Corruption perception index 2016*. Retrieved July 6, 2017, from https://www.transparency.org/news/feature/corruption_perceptions_index_2016
- Utikal, V., & Fischbacher, U. (2013). Disadvantageous lies in individual decisions. *Journal of Economic Behavior & Organization*, 85, 108–111. <http://dx.doi.org/10.1016/j.jebo.2012.11.011>
- Van Zant, A. B., & Kray, L. J. (2014). “I can’t lie to your face”. Minimal face-to-face interaction promotes honesty. *Journal of Experimental Social Psychology*, 55, 234–238. <http://dx.doi.org/10.1016/j.jesp.2014.07.014>
- Vetter, S. (2012). *Empirical studies of individual behavior. Cheating, corruption, and insurance choice* (Doctoral dissertation). Germany: Ludwig Maximilian University of Munich. Retrieved from <https://edoc.ub.uni-muenchen.de/15005/>
- Vevea, J. L., & Hedges, L. V. (1995). A general linear model for estimating effect size in the presence of publication bias. *Psychometrika*, 60, 419–435. <http://dx.doi.org/10.1007/BF02294384>
- Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software*, 36, 1–48. <http://dx.doi.org/10.18637/jss.v036.i03>
- Wang, J., Rao, Y., & Houser, D. E. (2017). An experimental analysis of acquired impulse control among adult humans intolerant to alcohol. *Proceedings of the National Academy of Sciences of the United States of America*, 114, 1299–1304. <http://dx.doi.org/10.1073/pnas.1610902114>
- Wang, L., & Murnighan, J. K. (2017). How much does honesty cost? Small bonuses can motivate ethical behavior. *Management Science*, 63, 2903–2914. <http://dx.doi.org/10.1287/mnsc.2016.2480>
- Welsh, D. T., Ellis, A. P. J., Christian, M. S., & Mai, K. M. (2014). Building a self-regulatory model of sleep deprivation and deception: The role of caffeine and social influence. *Journal of Applied Psychology*, 99, 1268–1277. <http://dx.doi.org/10.1037/a0036202>
- Wibral, M., Dohmen, T., Klingmüller, D., Weber, B., & Falk, A. (2012). Testosterone administration reduces lying in men. *PLoS ONE*, 7(10), e46774. <http://dx.doi.org/10.1371/journal.pone.0046774>
- Wiltermuth, S. S. (2011). Cheating more when the spoils are split. *Organizational Behavior and Human Decision Processes*, 115, 157–168. <http://dx.doi.org/10.1016/j.obhdp.2010.10.001>
- Winterich, K. P., Mittal, V., & Morales, A. C. (2014). Protect thyself: How affective self-protection increases self-interested, unethical behavior. *Organizational Behavior and Human Decision Processes*, 125, 151–161. <http://dx.doi.org/10.1016/j.obhdp.2014.07.004>
- Yaniv, G., & Siniver, E. (2016). The (honest) truth about rational dishonesty. *Journal of Economic Psychology*, 53, 131–140. <http://dx.doi.org/10.1016/j.joep.2016.01.002>
- Zettler, I., Hilbig, B. E., Moshagen, M., & de Vries, R. E. (2015). Dishonest responding or true virtue? A behavioral test of impression. *Personality and Individual Differences*, 81, 107–111. <http://dx.doi.org/10.1016/j.paid.2014.10.007>
- Zhang, H., Shi, Y., Zhou, Z. E., Ma, H., & Tang, H. (2018). Good people do bad things: How anxiety promotes unethical behavior through intuitive and automatic processing. *Current Psychology*. Advance online publication. <http://dx.doi.org/10.1007/s12144-018-9789-7>
- Zhong, C.-B. (2011). The ethical dangers of deliberative decision making. *Administrative Science Quarterly*, 56, 1–25. <http://dx.doi.org/10.2189/asqu.2011.56.1.001>
- Zhong, C.-B., Bohns, V. K., & Gino, F. (2010). Good lamps are the best police: Darkness increases dishonesty and self-interested behavior. *Psychological Science*, 21, 311–314. <http://dx.doi.org/10.1177/0956797609360754>

Appendix A

Inclusion Criteria and Integrated Experiments

This section details the article-selection process. First, to exclude possible learning effects within experiments, we integrated only *one-shot* experiments in which participants had a single opportunity to behave dishonestly. Second, we integrated only experiments that guaranteed full *anonymity* to participants and any partner(s). Revealing the *participant's* identity may elicit reputational concerns, such as honor and shame. Revealing the *assigned partner's* identity may elicit distinct other-regarding preferences, such as different preferences for (mis)reporting to women and men (e.g., Van Zant & Kray, 2014). Third, to rule out fear of explicit punishment as a possible motive for honest reporting, we integrated only experiments in which dishonest actions could not be *sanctioned*. Fourth, the outcome that would be reported by a completely honest sample had to be *unambiguous*. For example, matrix tasks must have measured the number of matrices that were solved (by either a control group or the participants themselves). Fifth, we excluded experiments in which dishonest behavior was

not directly *incentivized*. For example, experiments in which only third parties, but not the participants themselves, profited from dishonest behavior were excluded.

If the above criteria were fulfilled, we contacted the authors to request the primary data. We wrote at least three emails to the corresponding author. If authors did not reply, could not share the data (for technical reasons), or refused to share the data, we retrieved the necessary data points from the figures and summary statistics reported in the article. Three articles had to be excluded from further analyses because the authors did not share the primary data and figures or summary statistics were not provided in a codable manner (Gino & Mogilner, 2014; Rigdon & D'Esterre, 2017; Zhang, Shi, Zhou, Ma, & Tang, 2018).

The following tables list all integrated experiments in chronological order: sender–receiver games (Table A1), coin-flip tasks (Table A2), die-roll tasks (Table A3), and matrix tasks (Table A4). Each row represents an experimental treatment.

Table A1
Integrated Sender–Receiver Games

Study: Condition	<i>n</i>	<i>M_r</i>	Data
Gneezy, 2005: Treatment 1	75	36%	n.r.
Gneezy, 2005: Treatment 2	75	17%	n.r.
Gneezy, 2005: Treatment 3	75	52%	n.r.
Dreber & Johannesson, 2008	156	47%	yes
Cohen, Gunia, Kim-Jun, & Murnighan, 2009: Individual, certain	46	48%	yes
Cohen et al., 2009: Group, certain	38	71%	yes
Cohen et al., 2009: Individual, uncertain	37	32%	yes
Cohen et al., 2009: Group, uncertain	31	19%	yes
Hurkens & Kartik, 2009: Treatment 4	58	38%	yes
Hurkens & Kartik, 2009: Treatment 5	32	47%	yes
Sutter, 2009: Individual, Treatment 1	96	44%	yes
Sutter, 2009: Individual, Treatment 2	96	35%	yes
Sutter, 2009: Individual, Treatment 3	93	59%	yes
Sutter, 2009: Group, Treatment 1	22	23%	yes
Sutter, 2009: Group, Treatment 2	22	23%	yes
Sutter, 2009: Group, Treatment 3	24	25%	yes
Rode, 2010: Study 1, cooperative	32	81%	yes
Rode, 2010: Study 1, competitive	32	75%	yes
Rode, 2010: Study 2, cooperative	54	70%	yes
Rode, 2010: Study 2, competitive	54	70%	yes
Rode, 2010: Study 3, cooperative	27	81%	yes
Rode, 2010: Study 4, competitive	27	78%	yes
Cohen, Wolf, Panter, & Insko, 2011: Study 2	72	32%	yes

(Appendices continue)

Table A1 (*continued*)

Study: Condition	<i>n</i>	<i>M_r</i>	Data
Zhong, 2011: Study 1, deliberative	22	68%	yes
Zhong, 2011: Study 1, intuitive	22	36%	yes
Zhong, 2011: Study 2, deliberative	21	62%	yes
Zhong, 2011: Study 2, intuitive	20	30%	yes
Burks & Krupka, 2012	27	26%	n.d.
Childs, 2012b: Gain frame	49	53%	yes
Childs, 2012b: Loss frame	47	62%	yes
Erat & Gneezy, 2012: Study 1, T[1, 10]	101	49%	n.r.
Erat & Gneezy, 2012: Study 1, T[10, 10]	102	65%	n.r.
Erat & Gneezy, 2012: Study 1, T[1, -5]	104	37%	n.r.
Erat & Gneezy, 2012: Study 1, T[10, 0]	109	52%	n.r.
Gunia, Wang, Huang, Wang, & Murnighan, 2012: contemplation	30	13%	yes
Gunia et al., 2012: Immediate	34	44%	yes
Gunia et al., 2012: Moral	25	20%	yes
Gunia et al., 2012: Control	29	31%	yes
Gunia et al., 2012: Self-interested	28	50%	yes
Hershfield, Cohen, & Thompson, 2012: Study 3	27	56%	yes
Vetter, 2012: Study 2, control	48	44%	n.d.
Vetter, 2012: Study 2, responsibility	46	48%	n.d.
Angelova & Regner, 2013: Study 1, obligatory €1, with payment	32	56%	yes
Angelova & Regner, 2013: Study 1, obligatory €1, without payment	32	75%	yes
Angelova & Regner, 2013: Study 1, obligatory €2, with payment	32	66%	yes
Angelova & Regner, 2013: Study 1, obligatory €2, without payment	32	81%	yes
Angelova & Regner, 2013: Study 1, voluntary €1, with payment	31	81%	yes
Angelova & Regner, 2013: Study 1, voluntary €1, without payment	31	94%	yes
Angelova & Regner, 2013: Study 1, voluntary €2, with payment	32	59%	yes
Angelova & Regner, 2013: Study 1, voluntary €2, without payment	32	78%	yes
Aoki, Akai, & Onoshiro, 2013: A100-S	78	41%	yes
Aoki et al., 2013: A100-NS	31	32%	yes
Aoki et al., 2013: A1000-S	70	49%	yes
Aoki et al., 2013: A1000-NS	27	26%	yes
Cappelen, Sørensen, & Tungodden, 2013: Base	68	69%	yes
Cappelen et al., 2013: Market	65	72%	yes
Cappelen et al., 2013: Intuition	69	42%	yes
Cappelen et al., 2013: Personal	67	55%	yes
Childs, 2013: Gain frame	48	58%	yes
Childs, 2013: Loss frame	49	41%	yes
Erat, 2013: Treatment T[-2]	132	67%	n.r.
Erat, 2013: Treatment T[-6]	131	64%	n.r.
Gu, Zhong, & Page-Gould, 2013: Study 2, normal	31	58%	yes
Gu et al., 2013: Study 2, fast	32	31%	yes
Gu et al., 2013: Study 3, normal	31	71%	yes
Gu et al., 2013: Study 3, fast	36	47%	yes
Gu et al., 2013: Study 4, normal, decision-making	45	56%	yes
Gu et al., 2013: Study 4, normal, intuitive	46	63%	yes
Gu et al., 2013: Study 4, fast, decision-making	37	62%	yes
Gu et al., 2013: Study 4, fast, intuitive	40	38%	yes
Gylfason, Arnardottir, & Kristinsson, 2013	184	44%	yes
Innes & Mitra, 2013: Arizona, control	97	41%	yes
Innes & Mitra, 2013: Arizona, 15% untruthful	25	36%	yes
Innes & Mitra, 2013: Arizona, 40% untruthful	26	46%	yes
Innes & Mitra, 2013: Arizona, 60% untruthful	33	46%	yes
Innes & Mitra, 2013: Arizona, 85% untruthful	52	81%	yes
Innes & Mitra, 2013: California, control	26	42%	yes
Innes & Mitra, 2013: California, 0-2 untruthful	27	37%	yes
Innes & Mitra, 2013: California, 3-5 untruthful	52	71%	yes
Innes & Mitra, 2013: India, control	54	56%	yes
Innes & Mitra, 2013: India, 15% untruthful	39	31%	yes
Innes & Mitra, 2013: India, 85% untruthful	38	71%	yes
Innes & Mitra, 2013: India, control	54	56%	yes

(Appendices continue)

Table A1 (continued)

Study: Condition	<i>n</i>	<i>M_r</i>	Data
Innes & Mitra, 2013: India, 0–2 untruthful	46	37%	yes
Innes & Mitra, 2013: India, 3–5 untruthful	37	81%	yes
Kouchaki, Smith-Crowe, Brief, & Sousa, 2013: Study 3, money	46	46%	yes
Kouchaki et al., 2013: Study 3, control	45	22%	yes
López-Pérez & Spiegelman, 2013	30	60%	yes
Kouchaki & Smith, 2014: Study 3, afternoon	51	65%	yes
Kouchaki & Smith, 2014: Study 3, morning	51	43%	yes
Ter Meer, 2014: No feedback, piece rate	21	52%	n.r.
Ter Meer, 2014: No feedback, revenue sharing	21	68%	n.r.
Ter Meer, 2014: No feedback, tournament incentive	21	87%	n.r.
Ter Meer, 2014: Feedback, piece rate	24	78%	n.r.
Ter Meer, 2014: Feedback, revenue sharing	24	67%	n.r.
Ter Meer, 2014: Feedback, tournament incentive	24	67%	n.r.
Welsh, Ellis, Christian, & Mai, 2014: deprivation, influence, caffeine	29	52%	yes
Welsh et al., 2014: No deprivation, influence, caffeine	31	55%	yes
Welsh et al., 2014: Deprivation, no influence, caffeine	31	52%	yes
Welsh et al., 2014: No deprivation, no influence, caffeine	25	48%	yes
Welsh et al., 2014: Deprivation, influence, no caffeine	28	79%	yes
Welsh et al., 2014: No deprivation, influence, no caffeine	28	50%	yes
Welsh et al., 2014: Deprivation, no influence, no caffeine	25	32%	yes
Welsh et al., 2014: No deprivation, no influence, no caffeine	30	47%	yes
Winterich, Mittal, & Morales, 2014: Study 1b, neutral	25	35%	n.r.
Winterich et al., 2014: Study 1b, disgust	25	67%	n.r.
Winterich et al., 2014: Study 3, neutral, cleansing	32	43%	n.r.
Winterich et al., 2014: Study 3, neutral, no cleansing	32	30%	n.r.
Winterich et al., 2014: Study 3, disgust, no cleansing	32	35%	n.r.
Winterich et al., 2014: Study 3, disgust, no cleansing	32	53%	n.r.
Biziou-van-Pol et al., 2015: Pareto white lie	598	83%	yes
Lee, Im, Parmar, & Gino, 2015: Study 2	160	43%	yes
Mai, Ellis, & Welsh, 2015: Creativity activation	93	44%	yes
Mai et al., 2015: No creativity activation	85	41%	yes
Peeters, Vorsatz, & Walzl, 2015: Treatment SR	192	28%	yes
Gylfason, Halldorsson, & Kristinsson, 2016	143	36%	yes
Kilduff, Galinsky, Gallo, & Reade, 2016: Study 2, Michigan	26	46%	yes
Kilduff et al., 2016: Study 2, Berkeley	23	13%	yes
Kilduff et al., 2016: Study 2, Virginia	21	19%	yes
Roeser et al., 2016: Message task	195	22%	yes
Wang & Murnighan, 2017: Study 1, six-choice, baseline	20	50%	yes
Wang & Murnighan, 2017: Study 1, six-choice, exhortation	20	40%	yes
Wang & Murnighan, 2017: Study 1, six-choice, maximize profit	20	50%	yes
Wang & Murnighan, 2017: Study 1, six-choice, money for honesty	22	14%	yes
Wang & Murnighan, 2017: Study 1, six-choice, extra money only	20	40%	yes
Wang & Murnighan, 2017: Study 1, two-choice, baseline	17	47%	yes
Wang & Murnighan, 2017: Study 1, two-choice, exhortation	18	61%	yes
Wang & Murnighan, 2017: Study 1, two-choice, maximize profit	15	53%	yes
Wang & Murnighan, 2017: Study 1, two-choice, money for honesty	22	14%	yes
Wang & Murnighan, 2017: Study 1, two-choice, extra money only	16	69%	yes
Wang & Murnighan, 2017: Study 2, baseline no. 1	23	61%	yes
Wang & Murnighan, 2017: Study 2, baseline no. 2	23	70%	yes
Wang & Murnighan, 2017: Study 2, money for honesty	24	29%	yes
Wang & Murnighan, 2017: Study 2, exhortation no. 1	26	58%	yes
Wang & Murnighan, 2017: Study 2, exhortation no. 2	24	54%	yes
Wang & Murnighan, 2017: Study 3, baseline, \$4	42	50%	yes
Wang & Murnighan, 2017: Study 3, baseline, \$8	36	61%	yes
Wang & Murnighan, 2017: Study 3, baseline, \$12	47	51%	yes
Wang & Murnighan, 2017: Study 3, baseline, \$16	30	53%	yes
Wang & Murnighan, 2017: Study 3, baseline, \$20	39	44%	yes
Wang & Murnighan, 2017: Study 3, money for honesty, \$4	43	28%	yes
Wang & Murnighan, 2017: Study 3, money for honesty, \$8	43	35%	yes
Wang & Murnighan, 2017: Study 3, money for honesty, \$12	40	33%	yes
Wang & Murnighan, 2017: Study 3, money for honesty, \$16	44	27%	yes

(Appendices continue)

Table A1 (*continued*)

Study: Condition	<i>n</i>	<i>M_r</i>	Data
Wang & Murnighan, 2017: Study 3, money for honesty, \$20	36	44%	yes
Wang & Murnighan, 2017: Study 4, baseline	23	57%	yes
Wang & Murnighan, 2017: Study 4, money for honesty, \$1.00	21	38%	yes
Wang & Murnighan, 2017: Study 4, money for honesty, \$.75	21	14%	yes
Wang & Murnighan, 2017: Study 4, money for honesty, \$.50	20	45%	yes
Wang & Murnighan, 2017: Study 4, money for honesty, \$.25	27	44%	yes
Capraro, 2017: Time delay	516	44%	yes
Capraro, 2017: Time pressure	497	57%	yes
Kerschbamer, Neururer, & Gruber, 2017	118	75%	yes
Leibbrandt, Maitra, & Neelim, 2017: Hi1	15	40%	n.r.
Leibbrandt et al., 2017: Hi2	15	67%	n.r.
Leibbrandt et al., 2017: Mo1	30	20%	n.r.
Leibbrandt et al., 2017: Mo2	30	53%	n.r.
Leibbrandt et al., 2017: Mo3	31	29%	n.r.
Leibbrandt et al., 2017: Si1	31	36%	n.r.
Leibbrandt et al., 2017: Si2	31	26%	n.r.
Leibbrandt et al., 2017: Si3	31	42%	n.r.
Gawn & Innes, 2018	266	41%	n.r.
Zhang, Shi, Zhou, Ma, & Tang, 2018: Study 1, anxiety	32	83%	n.r.
Zhang et al., 2018: Study 1, neutral	32	83%	n.r.

Note. *n* = number of participants; *M_r* = standardized report; data = primary data shared (with yes = data set shared; n.r. = no response from the authors; n.d. = responded but data set inaccessible).

Table A2

Integrated Coin-Flip Tasks

Study: Condition	<i>n</i>	<i>M_r</i>	Data
Buccioli & Piovesan, 2011: Control	89	71%	yes
Buccioli & Piovesan, 2011: Request	93	38%	yes
Houser, Vetter, & Winter, 2012: Receivers who earned €0	96	65%	n.d.
Houser et al., 2012: Receivers who earned €2	75	41%	n.d.
Houser et al., 2012: Receivers who earned ≥€4	80	50%	n.d.
Houser et al., 2012: Proposers	251	45%	n.d.
Houser et al., 2012: No intentions treatment, €0	92	41%	n.d.
Houser et al., 2012: No intentions treatment, €2	71	52%	n.d.
Houser et al., 2012: No intentions treatment, ≥€4	75	36%	n.d.
Fosgaard, Hansen, & Piovesan, 2013: 5 wins, handwritten	53	32%	yes
Fosgaard et al., 2013: 5 wins, preprinted	53	28%	yes
Fosgaard et al., 2013: 10 wins, handwritten	51	69%	yes
Fosgaard et al., 2013: 10 wins, preprinted	52	42%	yes
Hilbig & Hessler, 2013: Target number 1	127	10%	yes
Hilbig & Hessler, 2013: Target number 2	127	22%	yes
Hilbig & Hessler, 2013: Target number 3	128	33%	yes
Hilbig & Hessler, 2013: Target number 4	128	34%	yes
Hilbig & Hessler, 2013: Target number 5	128	32%	yes
Hilbig & Hessler, 2013: Target number 6	127	9%	yes
Ploner & Regner, 2013: Hidden roll, DG, philanthropy	60	87%	yes
Ploner & Regner, 2013: Hidden roll, DG, VCG punishment	32	69%	yes
Ploner & Regner, 2013: Hidden roll, DG, stand-alone	96	67%	yes
Ploner & Regner, 2013: Hidden roll, bonus, philanthropy	64	66%	yes
Ploner & Regner, 2013: Hidden roll, bonus, VCG punishment	32	88%	yes

(Appendices continue)

Table A2 (continued)

Study: Condition	<i>n</i>	<i>M_r</i>	Data
Ploner & Regner, 2013: Hidden roll, bonus, stand-alone	32	69%	yes
Abeler, Becker, & Falk, 2014: 1-coin, telephone	658	-11%	yes
Gino & Wiltermuth, 2014: Study 4	178	24%	n.d.
Muñoz-Izquierdo, de Liaño, Rin-Sánchez, & Pascual-Ezama, 2014: No penalty	90	27%	yes
Muñoz-Izquierdo et al., 2014: Penalty	90	40%	yes
Muñoz-Izquierdo et al., 2014: Altruistic penalty	90	16%	yes
Winterich et al., 2014: Study 1a, neutral	98	4%	n.r.
Winterich et al., 2014: Study 1a, disgust	98	26%	n.r.
Dubois, Rucker, & Galinsky, 2015: Self-beneficial condition	75	23%	yes
Hilbig & Zettler, 2015: Study 2	88	34%	yes
Hilbig & Zettler, 2015: Study 3	185	15%	yes
Hilbig & Zettler, 2015: Study 4, concealed game	50	100%	yes
Hilbig & Zettler, 2015: Study 6, standard condition	107	34%	yes
Hilbig & Zettler, 2015: Study 6, common goods condition	101	25%	yes
Pascual-Ezama et al., 2015: SRT, Austria	30	-20%	yes
Pascual-Ezama et al., 2015: SRT, Belgium	30	13%	yes
Pascual-Ezama et al., 2015: SRT, Colombia	30	13%	yes
Pascual-Ezama et al., 2015: SRT, Denmark	30	40%	yes
Pascual-Ezama et al., 2015: SRT, Finland	30	27%	yes
Pascual-Ezama et al., 2015: SRT, Germany	30	47%	yes
Pascual-Ezama et al., 2015: SRT, Greece	30	20%	yes
Pascual-Ezama et al., 2015: SRT, India	30	-7%	yes
Pascual-Ezama et al., 2015: SRT, Indonesia	30	20%	yes
Pascual-Ezama et al., 2015: SRT, Italy	30	20%	yes
Pascual-Ezama et al., 2015: SRT, Japan	30	13%	yes
Pascual-Ezama et al., 2015: SRT, Netherlands	30	33%	yes
Pascual-Ezama et al., 2015: SRT, Spain	30	53%	yes
Pascual-Ezama et al., 2015: SRT, Turkey	30	7%	yes
Pascual-Ezama et al., 2015: SRT, UK	30	27%	yes
Pascual-Ezama et al., 2015: SRT, USA	30	40%	yes
Pascual-Ezama et al., 2015: WRT, Austria	30	53%	yes
Pascual-Ezama et al., 2015: WRT, Belgium	30	-7%	yes
Pascual-Ezama et al., 2015: WRT, Colombia	30	33%	yes
Pascual-Ezama et al., 2015: WRT, Denmark	30	-13%	yes
Pascual-Ezama et al., 2015: WRT, Finland	30	33%	yes
Pascual-Ezama et al., 2015: WRT, Germany	30	0%	yes
Pascual-Ezama et al., 2015: WRT, Greece	30	20%	yes
Pascual-Ezama et al., 2015: WRT, India	30	33%	yes
Pascual-Ezama et al., 2015: WRT, Indonesia	30	13%	yes
Pascual-Ezama et al., 2015: WRT, Italy	30	-7%	yes
Pascual-Ezama et al., 2015: WRT, Japan	30	20%	yes
Pascual-Ezama et al., 2015: WRT, Netherlands	30	20%	yes
Pascual-Ezama et al., 2015: WRT, Spain	30	34%	yes
Pascual-Ezama et al., 2015: WRT, Turkey	30	33%	yes
Pascual-Ezama et al., 2015: WRT, UK	30	-33%	yes
Pascual-Ezama et al., 2015: WRT, USA	30	7%	yes
Pascual-Ezama et al., 2015: VRT, Austria	30	20%	yes
Pascual-Ezama et al., 2015: VRT, Belgium	30	7%	yes
Pascual-Ezama et al., 2015: VRT, Colombia	30	-13%	yes
Pascual-Ezama et al., 2015: VRT, Denmark	30	-20%	yes
Pascual-Ezama et al., 2015: VRT, Finland	30	7%	yes
Pascual-Ezama et al., 2015: VRT, Germany	30	7%	yes
Pascual-Ezama et al., 2015: VRT, Greece	30	-27%	yes
Pascual-Ezama et al., 2015: VRT, India	30	0%	yes
Pascual-Ezama et al., 2015: VRT, Indonesia	30	-20%	yes
Pascual-Ezama et al., 2015: VRT, Italy	30	13%	yes
Pascual-Ezama et al., 2015: VRT, Japan	30	13%	yes
Pascual-Ezama et al., 2015: VRT, Netherlands	30	0%	yes
Pascual-Ezama et al., 2015: VRT, Spain	30	26%	yes
Pascual-Ezama et al., 2015: VRT, Turkey	30	13%	yes

(Appendices continue)

Table A2 (continued)

Study: Condition	<i>n</i>	<i>M_r</i>	Data
Pascual-Ezama et al., 2015: VRT, UK	30	40%	yes
Pascual-Ezama et al., 2015: VRT, USA	30	27%	yes
Zettler, Hilbig, Moshagen, & de Vries, 2015: Study 2	134	17%	yes
Conrads et al., 2016: Individual treatment	114	65%	yes
Conrads et al., 2016: Team treatment	67	40%	yes
Dieckmann, Grimm, Unfried, Utikal, & Valmasoni, 2016: Germany	203	34%	yes
Dieckmann et al., 2016: France	202	54%	yes
Dieckmann et al., 2016: Italy	202	49%	yes
Dieckmann et al., 2016: Netherlands	204	65%	yes
Dieckmann et al., 2016: Spain	204	45%	yes
Hilbig, Moshagen, & Zettler, 2016: BFI	311	34%	yes
Hilbig et al., 2016: IPIP Big Five Marker Scale	308	36%	yes
Hilbig et al., 2016: NEO-FFI	310	27%	yes
Houser, List, Piovesan, Samek, & Winter, 2016: Parent alone, prize for parent	61	10%	yes
Hugh-Jones, 2016: High treatment, Argentina	40	35%	yes
Hugh-Jones, 2016: High treatment, Brazil	52	38%	yes
Hugh-Jones, 2016: High treatment, Switzerland	40	20%	yes
Hugh-Jones, 2016: High treatment, China	50	76%	yes
Hugh-Jones, 2016: High treatment, Denmark	33	33%	yes
Hugh-Jones, 2016: High treatment, United Kingdom	42	14%	yes
Hugh-Jones, 2016: High treatment, Greece	58	38%	yes
Hugh-Jones, 2016: High treatment, India	51	53%	yes
Hugh-Jones, 2016: High treatment, Japan	50	56%	yes
Hugh-Jones, 2016: High treatment, South Korea	62	55%	yes
Hugh-Jones, 2016: High treatment, Portugal	54	11%	yes
Hugh-Jones, 2016: High treatment, Russian Federation	49	43%	yes
Hugh-Jones, 2016: High treatment, Turkey	38	47%	yes
Hugh-Jones, 2016: High treatment, USA	36	11%	yes
Hugh-Jones, 2016: High treatment, South Africa	47	-11%	yes
Hugh-Jones, 2016: Low treatment, Argentina	58	31%	yes
Hugh-Jones, 2016: Low treatment, Brazil	49	47%	yes
Hugh-Jones, 2016: Low treatment, Switzerland	33	27%	yes
Hugh-Jones, 2016: Low treatment, China	51	65%	yes
Hugh-Jones, 2016: Low treatment, Denmark	56	21%	yes
Hugh-Jones, 2016: Low treatment, United Kingdom	47	-6%	yes
Hugh-Jones, 2016: Low treatment, Greece	36	6%	yes
Hugh-Jones, 2016: Low treatment, India	48	63%	yes
Hugh-Jones, 2016: Low treatment, Japan	49	63%	yes
Hugh-Jones, 2016: Low treatment, South Korea	41	61%	yes
Hugh-Jones, 2016: Low treatment, Portugal	45	33%	yes
Hugh-Jones, 2016: Low treatment, Russian Federation	57	44%	yes
Hugh-Jones, 2016: Low treatment, Turkey	47	15%	yes
Hugh-Jones, 2016: Low treatment, USA	45	47%	yes
Hugh-Jones, 2016: Low treatment, South Africa	45	20%	yes
Kajackaite & Gneezy, 2017: Basic cheating game, \$1	75	10%	yes
Kajackaite & Gneezy, 2017: Basic cheating game, \$5	75	15%	yes
Kajackaite & Gneezy, 2017: Basic cheating game, \$20	75	20%	yes
Kajackaite & Gneezy, 2017: Basic cheating game, \$50	75	8%	yes
Kajackaite & Gneezy, 2017: Basic mind game, \$1	75	18%	yes
Kajackaite & Gneezy, 2017: Basic mind game, \$5	75	36%	yes
Kajackaite & Gneezy, 2017: Basic mind game, \$20	75	29%	yes
Kajackaite & Gneezy, 2017: Basic mind game, \$50	75	39%	yes
Kajackaite & Gneezy, 2017: Extended cheating game, \$1	75	18%	yes
Kajackaite & Gneezy, 2017: Extended cheating game, \$5	75	18%	yes
Kajackaite & Gneezy, 2017: Extended cheating game, \$20	75	18%	yes
Kajackaite & Gneezy, 2017: Extended cheating game, \$50	75	15%	yes
Kajackaite & Gneezy, 2017: Extended mind game, \$1	75	14%	yes
Kajackaite & Gneezy, 2017: Extended mind game, \$5	75	15%	yes
Kajackaite & Gneezy, 2017: Extended mind game, \$20	75	39%	yes
Kajackaite & Gneezy, 2017: Extended mind game, \$50	75	39%	yes

(Appendices continue)

Table A2 (continued)

Study: Condition	<i>n</i>	<i>M_r</i>	Data
Peer, Brandimarte, Samat, & Acquisti, 2017: Amazon Mechanical Turk sample	201	68%	yes
Peer et al., 2017: Crowdflower sample	221	38%	yes
Peer et al., 2017: Prolific Academic sample	214	45%	yes
Schindler & Pfattheicher, 2017: Study 2, gain frame	154	42%	yes
Schindler & Pfattheicher, 2017: Study 2, loss frame	146	66%	yes
Thielmann, Hilbig, Zettler, & Moshagen, 2017: study 2	152	33%	yes
Cagala, Glogowsky, & Rincke, 2018: Control	65	16%	n.s.
Cagala et al., 2018: Ethics	43	19%	n.s.
Cagala et al., 2018: Neutral	75	26%	n.s.
Hilbig & Zettler, 2018: Pro-self cheating	57	37%	yes
Klein, Thielmann, Hilbig, & Zettler, 2018: Pro-self cheating	103	37%	yes
Lohse, Simon, & Konrad, 2018: Contemplation	117	44%	yes
Lohse et al., 2018: Time pressure	188	29%	yes
Thielmann & Hilbig, 2018a: Study 1	183	34%	yes
Thielmann & Hilbig, 2018a: Study 2	200	27%	yes
Thielmann & Hilbig, 2018b: €5	110	26%	yes
Thielmann & Hilbig, 2018b: €10	117	26%	yes
Thielmann & Hilbig, 2018b: €20	116	30%	yes
Moshagen, Hilbig, & Zettler, 2018	929	10%	yes

Note. *n* = number of participants; *M_r* = standardized report; data = primary data shared (with yes = data set shared; n.r. = no response from the authors; n.d. = responded but data set inaccessible; n.s. = responded but refused to share).

Table A3

Integrated Die-Roll Tasks

Study: Condition	<i>n</i>	<i>M_r</i>	Data
Lammers, Stapel, & Galinsky, 2010: Study 1, high power	17	41%	yes
Lammers et al., 2010: Study 1, low power	18	19%	yes
Gino & Ariely, 2012: Study 3, control, low justification	36	4%	n.d.
Gino & Ariely, 2012: Study 3, control, high justification	36	42%	n.d.
Gino & Ariely, 2012: Study 3, creative mindset, high justification	36	53%	n.d.
Gino & Ariely, 2012: Study 3, creative mindset, high justification	36	57%	n.d.
Gino & Ariely, 2012: Study 4, control, low justification	40	-23%	n.d.
Gino & Ariely, 2012: Study 4, control, high justification	40	15%	n.d.
Gino & Ariely, 2012: Study 4, creative mindset, low justification	40	41%	n.d.
Gino & Ariely, 2012: Study 4, creative mindset, high justification	40	47%	n.d.
Shalvi, Dana, Handgraaf, & De Dreu, 2011: Single roll	62	19%	yes
Shalvi et al., 2011: Multiple roll	67	38%	yes
Shalvi, Handgraaf, & De Dreu, 2011: €3.50 exit	25	20%	yes
Shalvi et al., 2011: €2.50 exit	30	32%	yes
Piff, Stancato, Côté, Mendoza-Denton, & Keltner, 2012: Study 6	189	4%	yes
Shalvi, Eldar, & Bereby-Meyer, 2012: Study 1, low time pressure	38	15%	yes
Shalvi et al., 2012: Study 2, high time pressure	34	42%	yes
Shalvi et al., 2012: Study 1, low time pressure	33	-3%	yes
Shalvi et al., 2012: Study 2, high time pressure	39	35%	yes
Wibral, Dohmen, Klingmüller, Weber, & Falk, 2012: placebo	45	45%	yes
Wibral et al., 2012: Testosterone	46	-5%	yes
Conrads, Irlenbusch, Rilke, & Walkowitz, 2013: Individual	156	32%	yes
Conrads et al., 2013: Team	132	54%	yes
Conrads et al., 2013: Team-mixed, individual	130	34%	yes
Conrads et al., 2013: Team-mixed, team	136	45%	yes
Fischbacher & Föllmi-Heusi, 2013: Baselines	389	41%	yes

(Appendices continue)

Table A3 (continued)

Study: Condition	<i>n</i>	<i>M_r</i>	Data
Fischbacher & Föllmi-Heusi, 2013: High stakes	80	34%	yes
Fischbacher & Föllmi-Heusi, 2013: 4.9	125	38%	yes
Fischbacher & Föllmi-Heusi, 2013: Externality	78	27%	yes
Fischbacher & Föllmi-Heusi, 2013: Double anonymous	137	37%	yes
Gino, Krupka, & Weber, 2013: Mandatory regulation in prior task	30	17%	n.d.
Gino et al., 2013: No regulation in prior task	30	44%	n.d.
Gino et al., 2013: Voluntary regulation in prior task	30	48%	n.d.
Gravert, 2013: Random income	57	8%	yes
Shalvi & Leiser, 2013: Religious track	65	9%	yes
Shalvi & Leiser, 2013: Regular track	61	27%	yes
Utikal & Fischbacher, 2013: Students	19	54%	yes
Utikal & Fischbacher, 2013: Nuns	12	-33%	yes
Abeler et al., 2014: 4-coin, telephone	94	-6%	yes
Abeler et al., 2014: 4-coin-lab, telephone	170	32%	yes
Abeler et al., 2014: 4-coin-lab, click	180	39%	yes
Arbel, Bar-El, Siniver, & Tobol, 2014: Study 1	205	27%	yes
Arbel et al., 2014: Study 3, incentive to lie	194	27%	yes
Clot, Grolleau, & Ibanez, 2014: Control	49	3%	yes
Clot et al., 2014: Good deed	49	26%	yes
Conrads, Irlenbusch, Rilke, Schielke, & Walkowitz, 2014: Treatment 1	159	37%	yes
Conrads et al., 2014: Treatment 3	159	48%	yes
Conrads et al., 2014: Treatment 5	160	54%	yes
Gunia, Barnes, & Sah, 2014: Study 2, MTurk	99	21%	yes
Gunia et al., 2014: Study 2, students	43	26%	yes
Peer, Acquisti, & Shalvi, 2014: Study 1	2113	16%	yes
Peer et al., 2014: Study 3	719	29%	yes
Ruffle & Tobol, 2014	427	15%	yes
Chou, 2015: Study 1, e-signature	30	28%	n.r.
Chou, 2015: Study 1, handwritten signature	28	11%	n.r.
Conrads & Lotz, 2015: Face-to-face	60	37%	yes
Conrads & Lotz, 2015: Phone	60	38%	yes
Conrads & Lotz, 2015: Computerized, lab	60	36%	yes
Conrads & Lotz, 2015: Computerized, remote	66	43%	yes
Jacobsen & Piovesan, 2016: Baseline	50	1%	yes
Jacobsen & Piovesan, 2016: Tax framing	50	18%	yes
Jacobsen & Piovesan, 2016: Explanation	49	-4%	yes
Muehlheusser, Roider, & Wallmeier, 2015: Individual	108	39%	yes
Muehlheusser et al., 2015: Team	60	39%	yes
Cadsby, Du, & Song, 2016: Die-roll self	90	34%	yes
Dai, Galeotti, & Villeval, 2018: Station	244	43%	yes
Dai, Galeotti, & Villeval, 2018: Fine collection office	35	26%	yes
Gächter & Schulz, 2016: Austria	66	30%	yes
Gächter & Schulz, 2016: China, high stakes	138	41%	yes
Gächter & Schulz, 2016: China, low stakes	99	43%	yes
Gächter & Schulz, 2016: Colombia	104	35%	yes
Gächter & Schulz, 2016: Czech Republic	77	35%	yes
Gächter & Schulz, 2016: Georgia	97	36%	yes
Gächter & Schulz, 2016: Germany	69	20%	yes
Gächter & Schulz, 2016: Guatemala	193	31%	yes
Gächter & Schulz, 2016: Indonesia	76	35%	yes
Gächter & Schulz, 2016: Italy	82	23%	yes
Gächter & Schulz, 2016: Kenya	92	37%	yes
Gächter & Schulz, 2016: Lithuania	71	18%	yes
Gächter & Schulz, 2016: Malaysia	64	29%	yes
Gächter & Schulz, 2016: Morocco	138	56%	yes
Gächter & Schulz, 2016: Netherlands	84	30%	yes
Gächter & Schulz, 2016: Poland	110	38%	yes
Gächter & Schulz, 2016: Slovakia	87	25%	yes
Gächter & Schulz, 2016: South Africa	92	29%	yes

(Appendices continue)

Table A3 (continued)

Study: Condition	<i>n</i>	<i>M_r</i>	Data
Gächter & Schulz, 2016: Spain	54	31%	yes
Gächter & Schulz, 2016: Sweden	82	19%	yes
Gächter & Schulz, 2016: Tanzania	140	58%	yes
Gächter & Schulz, 2016: Turkey	244	39%	yes
Gächter & Schulz, 2016: United Kingdom	197	19%	yes
Gächter & Schulz, 2016: Vietnam	112	38%	yes
Schurr & Ritov, 2016: Control	23	3%	yes
Schurr & Ritov, 2016: Study 1, winners	20	35%	yes
Schurr & Ritov, 2016: Study 1, losers	23	-13%	yes
Schurr & Ritov, 2016: Study 2, winners	19	38%	yes
Schurr & Ritov, 2016: Study 2, losers	19	3%	yes
Schurr & Ritov, 2016: Study 3a, winners	29	-20%	yes
Schurr & Ritov, 2016: Study 3a, losers	22	6%	yes
Schurr & Ritov, 2016: Study 3b, winners	23	5%	yes
Schurr & Ritov, 2016: Study 3b, losers	21	22%	yes
Bassarak et al., 2017: Observing, other's die	60	37%	yes
Bassarak et al., 2017: Observing, own die	61	29%	yes
Bassarak et al., 2017: Producing & observing	61	34%	yes
Kajackaite, 2018: Random draw	91	49%	yes
Korbel, 2017: Endogenous group	50	35%	yes
Korbel, 2017: Exogenous group	59	27%	yes
Korbel, 2017: Individual	117	17%	yes
Ruffle & Tobol, 2017: Follow-up experiment	156	27%	yes
Schindler & Pfattheicher, 2017: Study 1, gain frame	47	-5%	yes
Schindler & Pfattheicher, 2017: Study 1, loss frame	39	2%	yes
Wang, Rao, & Houser, 2017	477	27%	yes
Andersen, Gneezy, Kajackaite, & Marx, 2018: Delay, high stakes	40	66%	yes
Andersen et al., 2018: Delay, low stakes	67	66%	yes
Andersen et al., 2018: No delay, high stakes	41	61%	yes
Andersen et al., 2018: No delay, high stakes	65	81%	yes
Crone & Levy, 2018: Study 2	225	20%	yes
Crone & Levy, 2018: Study 3	301	22%	yes
Crone & Levy, 2018: Study 4, anti-free-will condition	95	20%	yes
Crone & Levy, 2018: Study 4, control	105	20%	yes
Duncan & Li, 2018: Self-report	203	30%	yes
Ezquerra, Kolev, & Rodriguez-Lara, 2018: Gain frame	89	40%	yes
Ezquerra et al., 2018: Loss frame	84	38%	yes
Gneezy, Kajackaite, & Sobel, 2018: 100-states, nonobserved	104	58%	yes
Gneezy et al., 2018: Basic, nonobserved	103	51%	yes
Gneezy et al., 2018: Low probability, nonobserved	107	41%	yes
Gneezy et al., 2018: Numbers, mixed	110	34%	yes
Gneezy et al., 2018: Numbers	390	34%	yes
Gneezy et al., 2018: Words	102	42%	yes
Kleinlogel, Dietz, & Antonakis, 2018: Study 1	195	16%	yes

Note. *n* = number of participants; *M_r* = standardized report; data = primary data shared (with yes = data set shared; n.r. = no response from the authors; n.d. = responded but data set inaccessible).

(Appendices continue)

Table A4
Integrated Matrix Tasks

Study: Condition	<i>n</i>	<i>M_t</i>	Data
Mazar, Amir, & Ariely, 2008: Study 1, Ten Commandments	116	-11%	yes
Mazar et al., 2008: Study 1, 10 books	113	7%	yes
Mazar et al., 2008: \$.50, control vs. recycle	61	17%	yes
Mazar et al., 2008: \$2, control vs. recycle	77	10%	yes
Mazar et al., 2008: Study 3, control vs. recycle	300	16%	yes
Mazar et al., 2008: Study 4	44	16%	yes
Mazar et al., 2008: Study 5, four matrices solved	52	7%	yes
Mazar et al., 2008: Study 5, eight matrices solved	56	8%	yes
Rhine, 2008: USA, no cheating vs. cheating to self	85	2%	n.r.
Rhine, 2008: China, no cheating vs. cheating to self	98	1%	n.r.
Gino, Ayal, & Ariely, 2009: Study 1, control vs. shredder	76	39%	n.d.
Gino et al., 2009: Study 2, control vs. shredder	61	40%	n.d.
Mead, Baumeister, Gino, Schweitzer, & Ariely, 2009: Study 1, no depletion	71	3%	n.d.
Mead et al., 2009: Study 1, depletion	62	18%	n.d.
Gino, Norton, & Ariely, 2010: Study 1a, authentic sunglasses	43	6%	n.d.
Gino et al., 2010: Study 1a, counterfeit sunglasses	42	28%	n.d.
Gino et al., 2010: Study 1b, authentic sunglasses	46	6%	n.d.
Gino et al., 2010: Study 1b, counterfeit sunglasses	45	21%	n.d.
Gino et al., 2010: Study 3, control	33	12%	n.d.
Gino et al., 2010: Study 3, authentic sunglasses	33	9%	n.d.
Gino et al., 2010: Study 3, counterfeit sunglasses	34	28%	n.d.
Zhong, Bohns, & Gino, 2010: Study 1, control room	42	6%	n.d.
Zhong et al., 2010: Study 1, dim room	42	33%	n.d.
Gino & Ariely, 2012: Study 2, control	56	9%	n.d.
Gino & Ariely, 2012: Study 2, creative mindset	56	21%	n.d.
Gino & Margolis, 2011: Study 3, prevention focus	41	8%	n.d.
Gino & Margolis, 2011: Study 3, aspiration, promotion focus	34	38%	n.d.
Gino & Margolis, 2011: Study 3, aspiration, prevention focus	34	12%	n.d.
Gino & Margolis, 2011: Study 3, compliance, promotion focus	34	22%	n.d.
Gino & Margolis, 2011: Study 3, compliance, prevention focus	34	4%	n.d.
Gino & Margolis, 2011: Study 3, promotion focus	41	30%	n.d.
Gino et al., 2011: Study 1, no depletion	51	8%	n.d.
Gino et al., 2011: Study 1, depletion	50	20%	n.d.
Gino et al., 2011: Study 2, no depletion	49	6%	n.d.
Gino et al., 2011: Study 2, depletion	48	18%	n.d.
Shu, Gino, & Bazerman, 2011: Study 3, no honor code	70	44%	n.r.
Shu et al., 2011: Study 3, honor code	70	17%	n.r.
Shu et al., 2011: Study 4, read honor code	22	22%	n.r.
Shu et al., 2011: Study 4, signed honor code	22	4%	n.r.
Shu et al., 2011: Study 4, control	23	44%	n.r.
Friesen & Gangadharan, 2012	114	11%	yes
Gino & Galinsky, 2012: Study 3, shared attributes	41	30%	n.d.
Gino & Galinsky, 2012: Study 3, control	41	11%	n.d.
Shu & Gino, 2012: Study 1	56	13%	n.r.
Shu & Gino, 2012: Study 2	78	12%	n.r.
Shu, Mazar, Gino, Ariely, & Bazerman, 2012: Study 1, top	35	8%	n.r.
Shu et al., 2012: Study 1, bottom	33	36%	n.r.
Shu et al., 2012: Study 1, control	33	24%	n.r.
Shu et al., 2012: Study 2, top	30	15%	n.r.
Shu et al., 2012: Study 2, bottom	30	31%	n.r.
Gamliel & Peer, 2013: Control vs. shredder	68	12%	yes
Gino, Ayal, & Ariely, 2013: Study 1, individual	64	30%	n.d.
Gino et al., 2013: Study 1, dyad	64	52%	n.d.
Gino et al., 2013: Study 1, group	64	68%	n.d.
Gino et al., 2013: Study 2, self only, high payoff	36	11%	n.d.
Gino et al., 2013: Study 2, self only, low payoff	35	10%	n.d.
Gino et al., 2013: Study 2, self and other payoff	36	27%	n.d.

(Appendices continue)

Table A4 (continued)

Study: Condition	<i>n</i>	<i>M_r</i>	Data
Gino et al., 2013: Study 1, no regulation vs. mandatory	60	25%	n.d.
Gravert, 2013: Performance income	48	41%	yes
Gunia et al., 2014: Study 1	48	5%	yes
Kouchaki, Gino, & Jami, 2014: Study 3, light backpack	35	15%	yes
Kouchaki et al., 2014: Study 3, heavy backpack	36	5%	yes
Rigdon & D'Esterre, 2014: Self-grading, noncompetitive	48	11%	yes
Rigdon & D'Esterre, 2014: Self-grading, competitive	52	9%	yes
Cai, Huang, Wu, & Kou, 2015: Study 1, eyes	66	9%	yes
Cai et al., 2015: Study 1, control	65	9%	yes
Faravelli, Friesen, & Gangadharan, 2015: Study 2, piece rate	119	25%	yes
Faravelli et al., 2015: Study 3, tournament rate	119	25%	yes
Faravelli et al., 2015: Study 4, piece rate	77	27%	yes
Faravelli et al., 2015: Study 4, tournament rate	41	36%	yes
Kouchaki & Wareham, 2015: Study 1, exclusion	20	38%	yes
Kouchaki & Wareham, 2015: Study 1, inclusion	18	21%	yes
Kouchaki & Wareham, 2015: Study 4, morning	27	15%	yes
Kouchaki & Wareham, 2015: Study 4, afternoon	21	28%	yes
Lee, Gino, Jin, Rice, & Josephs, 2015: Pilot study	82	16%	yes
Lee et al., 2015: Main study	117	17%	yes
Grolleau, Kocher, & Sutan, 2016: Gain frame	300	9%	yes
Grolleau et al., 2016: Loss frame	300	41%	yes
Hildreth, Gino, & Bazerman, 2016: Study 1a, loyalty	28	2%	yes
Hildreth et al., 2016: Study 1a, control	27	4%	yes
Hildreth et al., 2016: Study 1b, loyalty	33	2%	yes
Hildreth et al., 2016: Study 1b, control	30	13%	yes
Hildreth et al., 2016: Study 2b	88	2%	yes
Hildreth et al., 2016: Study 3a, loyalty	30	2%	yes
Hildreth et al., 2016: Study 3a, control	29	6%	yes
Hildreth et al., 2016: Study 3b, pledge	31	5%	yes
Hildreth et al., 2016: Study 3b, no pledge	39	2%	yes
Hildreth et al., 2016: Study 3b, control	36	7%	yes
Hildreth et al., 2016: Study 5a, loyalty, low competition	48	4%	yes
Hildreth et al., 2016: Study 5a, loyalty, high competition	51	14%	yes
Hildreth et al., 2016: Study 5a, control, low competition	52	13%	yes
Hildreth et al., 2016: Study 5a, control, high competition	55	17%	yes
Hildreth et al., 2016: Study 5b, loyalty, low competition	50	4%	yes
Hildreth et al., 2016: Study 5b, loyalty, high competition	53	12%	yes
Hildreth et al., 2016: Study 5b, control, low competition	55	10%	yes
Hildreth et al., 2016: Study 5b, control, high competition	50	11%	yes
Motro, Ordóñez, Pittarello, & Welsh, 2016: Anger	34	54%	yes
Motro et al., 2016: Guilt	35	6%	yes
Motro et al., 2016: Neutral	38	20%	yes
Kajackaite, 2018: Real effort control I vs. Real effort control II	95	34%	yes
Klimm, 2018	182	27%	yes

Note. *n* = number of participants; *M_r* = standardized report; data = primary data shared (with yes = data set shared; n.r. = no response from the authors; n.d. = responded but data set inaccessible).

(Appendices continue)

Appendix B

Classification of Experiments and Calculation of the Dishonesty Measures

Sender–Receiver Games

In sender–receiver games, a completely honest participant sample would always send the truthful message, where $t = t_{\min} = 0\%$, and a completely dishonest sample would send $t_{\max} = 100\%$ false message. The actual mean claim of the sample M_r is thus the percentage of false messages, where $M_r = m = M_{\text{liars}}$. A study with $m = 0\%$ honest messages converts into $M_r = 0\%$ and a study with $m = 100\%$ converts to $M_r = +100\%$. The standardized report for sender–receiver games can never take negative values. We address this methodological concern further below.

Coin-Flip Tasks

Coin-flip tasks covered all experiments in which (a) reporting was measured dichotomously—that is, involved a decision between claiming to have won a bonus and going empty handed—and (b) the “true” observed outcome had to be randomly generated. In most coin-flip tasks, participants tossed a single coin once and reported the outcome. In such situations, a completely honest sample would report a win about half of the time, $t = 50\%$. Reporting a win results in the highest possible claim, $t_{\max} = 100\%$, and reporting a loss, the lowest possible claim, $t_{\min} = 0\%$. A study with a mean claim of $m = 0\%$ would thus convert to $M_r = -100\%$; a study with $m = 50\%$ would convert to $M_r = 0\%$; and a study with $m = 100\%$ would convert to $M_r = +100\%$.

Some coin-flip tasks used multiple coin tosses to generate the “true” observed outcome. Here, participants earned a bonus for reporting a specific target number of wins (e.g., two heads in three tosses). If they did not report exactly the target number, they left empty handed (Hilbig & Zettler, 2015; Houser et al., 2012; Thielmann, Hilbig, Zettler, & Moshagen, 2017; Zettler, Hilbig, Moshagen, & de Vries, 2015). Other designs of coin-flip tasks randomly generated the “true” outcome via some form of dice game. Unlike die-roll tasks, such tasks enforce a dichotomous decision between reporting a win and a loss. For example, participants could win a flat bonus if they reported having rolled an even number (Ploner & Regner, 2013); if the sum of all pips reached a threshold (Conrads et al., 2016; Dubois, Rucker, & Galinsky, 2015); if the upward-facing side yielded a greater number than the downward-facing side (or vice versa; Peer et al., 2017); or if the die showed a specific target number, such as a 4 (Hilbig & Hessler, 2013; Hilbig

& Zettler, 2015). Alternatively, a wheel of fortune determined a winner (Lohse et al., 2018). The standardized report M_r and the rate of liars M_{liars} were always adapted to the specific experimental design. For example, a completely honest sample in a target-number dice game would report a win about one in six times; therefore, $t = 1/6$.

To estimate the rate of liars in coin-flip tasks, we assumed that participants would always claim a win if they tossed one. Accordingly, the rate of liars M_{liars} was estimated as

$$M_{\text{liars}} = 0\% \quad \text{if } m < t \quad (\text{B1})$$

$$M_{\text{liars}} = \frac{m-t}{1-t} \quad \text{if } m \geq t \quad (\text{B2})$$

where m is the actual report per study and t is the expected reported percentage of winning sides if participants were honest.

Estimating M_{liars} decreases the number of utilized observations per primary study because only a subset of participants tossed the losing side. We thus adjusted the sample size n per coin-flip task to be the estimated number of participants who tossed the losing side:

$$n = N \times (100\% - t) \quad (\text{B3})$$

where N is the total sample size per coin-flip task and t is the expected reported percentage of winning sides if participants were honest.

Die-Roll Tasks

Die-roll tasks included all experiments in which (a) the reported outcome was continuous and (b) the “true” observed outcome was either known (as in Piff, Stancato, Côté, Mendoza-Denton, & Keltner, 2012) or randomly generated. Most die-roll tasks used some form of dice game to generate the “true” outcome. For example, in a task with a six-sided die where each pip translates into \$1, the minimum report is $t_{\min} = 1$, the maximum report is $t_{\max} = 6$, and the expected report if everyone was honest is $t = 3.5$ —that is $(1 + 2 + 3 + 4 + 5 + 6)/6$, assuming the die is fair. In such a study, an average report of $m = 6$ would convert to $M_r = 100\%$, indicating that all participants who could cheat cheated to the maximal degree; an average report of $m = 3.5$ would convert to $M_r = 0\%$, indicating honest reporting; and an average report of $m = 1$ would convert to $M_r = -100\%$.

(Appendices continue)

Instead of dice, some experiments used repeated coin tosses, with participants being asked to report how many times they obtained the winning side (Abeler et al., 2014; Aydogan, Jobst, D'Ardenne, Müller, & Kocher, 2017; Conrads & Lotz, 2015). We classified these tasks as die-roll tasks because of their continuous measure of dishonesty. Notably, they were also one-shot tasks, because participants reported their observations only at a single instance. Most die-roll tasks converted pips into money by using either a complete linear payoff function (e.g., \$1 for each reported pip) or a linear payoff function that excluded the highest pip (e.g., \$1 for each reported pip, but reporting a 6 paid \$0). Only two die-roll tasks used alternative, nonlinear payoff functions (Ezquerro, Kolev, & Rodriguez-Lara, 2018: Reporting a 0 paid €0.00, a 1 paid €1.00, a 2 paid €1.50, a 3 paid €2.00, a 4 paid €2.50, a 5 paid €3.00, a 6 paid €3.50, a 7 paid €4.00, an 8 paid €4.50, and a 9 paid €5.00; Jacobsen & Piovesan, 2016: reporting a 1 paid DKK 10, a 2 paid DKK 20, a 3 paid DKK 30, a 4 paid DKK 35, a 5 paid DKK 40, and a 6 paid DKK 45; DKK = Danish krone). To calculate the standardized report for this study, we let m , t , and M_r refer to the reported outcomes (e.g., reporting a 5), not the associated payoffs (e.g., claiming DKK 40).

To calculate the rates of liars in die-roll tasks, we assumed that (a) all studies used fair dice and (b) participants would always claim at least the number they actually observed. Accordingly, we calculated the percentage of honest responses M_{honest} per primary experiment as:

$$M_{\text{honest}} = \frac{C(t_{\min})}{E(t_{\min})} \quad \text{if } E(t_{\min}) \leq C(t_{\min}) \quad (\text{B4})$$

$$M_{\text{honest}} = 0\% \quad \text{if } E(t_{\min}) > C(t_{\min}) \quad (\text{B5})$$

where $C(t_{\min})$ is the actual percentage of claims of the lowest paying option, and $E(t_{\min})$ is the expected percentage of claims of the lowest paying option if everyone was honest. For example, in a die-roll task with a six-sided die and a linear payoff function, the lowest option has a probability of $E(t_{\min}) = 1/6$. If, for example, only $C(t_{\min}) = 1/12$ of the participants report the lowest pip, then the estimated percentage of honest participants is $M_{\text{honest}} = 50\%$.

We assumed that the percentages of honest responses M_{honest} and dishonest responses M_{liars} per experiment would add up to 100%. Accordingly:

$$100\% - M_{\text{honest}} = M_{\text{liars}} \quad (\text{B6})$$

where M_{honest} is the percentage of honest responses and M_{liars} is the percentage of dishonest responses.

Two things are worth mentioning about *estimating* the rates of liars in die-roll tasks. First, M_{honest} is the lower bound for the percentage of honest reporting (assuming that participants would always claim at least the number they actually observed). This is because participants who actually obtained t_{\min} were maximally tempted to cheat (e.g., there might be participants who cheat when they observe the lowest outcome but would not cheat if they observe a medium outcome). Accordingly, the current estimation

of M_{liars} is an upper bound to the real rate of liars in the die-roll task. Second, estimating the rate of liars decreases the number of utilized observations per primary study because only a subset of participants obtained the lowest paying option, which is the basis for calculating M_{honest} . For all analyses that involve the rates of liars M_{liars} we thus adjusted the sample size n per die-roll task to be the estimated number of participants who obtained the lowest paying option per primary study:

$$n = N \times E(t_{\min}) \quad (\text{B7})$$

where N is the total sample size per die-roll task and $E(t_{\min})$ is the expected percentage of claims of the lowest paying option if everyone was honest.

Matrix Tasks

In all matrix tasks, participants were shown a series of 12 three-digit numbers (e.g., 6.41) and had to identify the two numbers that would add up to exactly 10. Honest participants would report the exact number of actually solved matrices. For matrix tasks that measured at the individual level, t reflects the “true” performance per experiment. For matrix tasks that measured at the aggregate level, we assumed that the control group (which could not possibly cheat) and the experimental group (which could cheat) solved an equal number of matrices. Hence, the “true” performance of the control group was assumed to be equal to the “true” performance of the experimental group, t . If, for example, on average $t = 5$ out of 20 matrices were solved, an honest participant sample would also claim an average of $m = 5$ matrices. The minimum is to claim no matrix was solved, $t_{\min} = 0$. The maximum is to claim all matrices were solved; for example, $t_{\max} = 20$ for a total of 20 matrices. In a study with 20 matrices and $t = 5$, an average claim of $m = 20$ would thus convert to $M_r = +100\%$, an average claim of $m = 5$ would convert to $M_r = 0\%$, and an average claim of $m = 0$ would convert to $M_r = -100\%$.

To calculate the rates of liars we limited the analysis to matrix tasks that measured dishonesty at the individual level and that either reported the rate of liars or for which we had the primary data. This held for only 77 of the 96 matrix tasks. For each experiment, we excluded (a) all participants who had solved all matrices and (b) those who reported having solved *fewer* matrices than was actually the case. These steps were necessary to exclude (a) participants who could not cheat to their own advantage because they had already solved all matrices (i.e., all subjects s_i whose performance p_i was equal to the maximum claim t_{\max}); and (b) participants who misreported in such a way that they earned less (i.e., all subjects s_i whose claim c_i was smaller than their performance p_i). We thus defined the eligible sample for calculating the rates of liars per primary study S^* as

$$S \supseteq S^* = \{s_i | p_i < t_{\max} \cap p_i \leq c_i\} \quad (\text{B8})$$

(Appendices continue)

where S denotes all observations per matrix task and S^* is the set of all subjects s_i whose performance p_i was smaller than the maximum possible claim t_{\max} and whose claims c_i were at least equal to the number of matrices they had solved.

Accordingly, the sample size n per primary study is the number of eligible observations, which was defined as

$$n = |S^*| \quad (\text{B9})$$

For each primary study, the percentage of honest responses M_{honest} was calculated as the fraction of the sample size n whose claims were equal to their true performance:

$$S^* \supseteq \text{honest} = \{s_i | c_i = p_i\} \quad (\text{B10})$$

$$M_{\text{honest}} = \frac{|\text{honest}|}{n} \quad (\text{B11})$$

where the honest reports per primary study, *honest*, are the reports of all participants s_i whose claims c_i equaled the number of matrices they had solved p_i .

It follows that the percentages of honest responses M_{honest} and dishonest responses M_{liars} per experiment add up to 100%. Accordingly,

$$100\% - M_{\text{honest}} = M_{\text{liars}} \quad (\text{B12})$$

where M_{honest} is the percentage of honest responses and M_{liars} is the percentage of dishonest responses.

Further Remarks

The calculation of the standardized report may cause biased estimates; that is, M_r and M_{liars} may systematically over- or underestimate the degree of dishonest responses per experiment. Biases can occur at two levels: *within* and *between* paradigms. Within-paradigm bias refers to variation in aspects of the experimental design that were specific to each paradigm. In die-roll tasks, for example, the “true” observed outcomes could be continuously uniform (e.g., the outcome of a one-shot die roll) or normally distributed (e.g., the outcome of two die rolls). We examine some potential reasons for within-paradigm bias in [Appendix C](#). Between-paradigm bias can occur as a result of systematic differences in one or more of the four experimental paradigms for methodological or theoretical reasons. Methodological reasons pertain to the way in which the standardized report was computed. For example, M_r can take negative values in die-roll, coin-flip, and matrix tasks, whereas in sender–receiver games $M_r = 0\%$ is the

minimum. To account for between-paradigm biases, we either analyzed the paradigms separately or used dummy variables to control for experimental paradigm. Theoretical reasons for between-paradigms biases cannot be addressed via statistical methods. For example, the four paradigms may in fact measure different facets of dishonest behavior: Standardized reports may represent the sending of false messages (sender–receiver games), the deviation from the expected value of a randomly generated figure (die-roll and coin-flip tasks), or one’s actual performance (matrix task).

For each study, we transformed the (provided or calculated) standard deviations of m into the units of the standardized report M_r :

$$s_r = \frac{s_m}{t - t_{\min}} \text{ if } m < t, \text{ and} \quad (\text{B13})$$

$$s_r = \frac{s_m}{t_{\max} - t} \text{ if } m \geq t \quad (\text{B14})$$

where s_r is the standard deviation of M_r and s_m is the standard deviation of m .

For all investigations we then estimated the standard error SE from the standard deviation s_m and the number of participants n per study:

$$SE = \frac{s_m}{\sqrt{n}} \quad (\text{B15})$$

Monetary Units

All monetary values were converted from local currency units to U.S. dollars using 2015 purchasing-power parity (PPP). The data for the PPP conversion were taken from the World Economic Outlook Database ([International Monetary Fund, 2015](#)). The 2015 PPP conversion rate for the Argentine peso was missing from the database; we therefore used its 2011 conversion rate. Two series of experiments used Lindt Lindor chocolate truffles as the incentive for reporting a win in coin-flip tasks ([Muñoz-Izquierdo et al., 2014](#); [Pascual-Ezama et al., 2015](#)). In March 2017, a pack of 60 Lindt Lindor chocolate truffles cost U.S. \$15.99 on Amazon.com. We therefore assumed that the incentive was worth \$0.27 (= \$15.99/60). In cases where the exact incentive size was unknown ([Gunia et al., 2014](#)) or other nonmonetary incentives were employed ([Hilbig & Zettler, 2015](#); [Hildreth et al., 2016, Study 2b](#); [Ruffle & Tobol, 2014](#)), we let the incentive size equal the mean incentive of the respective experimental paradigm.

(Appendices continue)

Appendix C

Supplemental Analyses With Experimental Paradigm-Specific Features

In this section, we analyze distinct experimental characteristics that are limited to specific experimental paradigms (e.g., variations in the setup practices of sender–receiver games). We also inspect the response distributions in experimental paradigms with continuous outcome measures (i.e., die-roll and matrix tasks).

Sender–Receiver Games

It is worth noting that the mean rate of 51% false messages may not represent the rate of intentionally misleading messages. In a seminal investigation, Sutter (2009) found that only about 70% of participants in the role of sender believed that their advice would be followed. Almost all of the 30% skeptical participants sent truthful messages, believing that their partners in the role of receivers would *not* follow their advice (sophisticated truth telling); only a minority sent false messages believing that their advice would *not* be followed. In our data set, the rate of advice following among receivers was indeed close to 70% ($k = 54$, $n = 3,321$, $M_{\text{follow}} = 67\%$). To reduce or even eliminate the possibility of sophisticated truth telling, researchers have commonly used two experimental modifications. Some experiments give more than two options for receivers to choose between. This setup makes it more difficult for receivers to find the option with the higher payoff by random guessing. The rationale is that the greater the number of options, the more receivers should trust senders. As a consequence, sophisticated truth tellers will reveal themselves by switching from sending true messages in games with two options to sending false messages in games with three and more options (Erat & Gneezy, 2012; Wang & Murnighan, 2017). A linear mixed effects model with random intercepts between experiments tested whether the number of options in sender–receiver games was associated with more lying. As predicted, every additional option raised the percentage of false messages by 2% (Table C1). An alternative method to reduce sophisticated truth telling is to inform participants that receivers have already decided to follow their advice (e.g., Cohen, Gunia, Kim-Jun, & Murnighan, 2009). However, we could not confirm that this approach was associated with fewer false messages (Table C1).

Adding the number of options to the regression model presented in Table 4 as a covariate did not qualitatively change the results, with one notable exception: The dummy variable online/telephone experiments reached conventional levels of significance (from $p = .061$ to $p = .047$; Table C2). The model fit improved marginally (from $R^2 = 37\%$ to $R^2 = 39\%$).

Table C1

Increasing the Number of Options Increased Misreporting in Sender–Receiver Games

Dependent variable	M_r
Intercept	42% (2.42)***
Number of options	2% (.58)***
Guaranteed implementation	1% (4.83)
Observations	$k = 165$ $n = 9,243$
Residual heterogeneity	$f^2 = 88\%$ $\tau^2 = .03$
Heterogeneity accounted for	$R^2 = 13\%$

Note. Linear mixed effects model with random intercepts between experiments. Unless denoted otherwise, values refer to regression estimates with standard errors in parentheses. M_r = Standardized report.

*** $p < .001$.

Die-Roll Tasks

In the following, we focus on die-roll tasks in which participants roll the die once. If participants were honest, rolling once would yield continuous uniform distributions of the reported score (see Figure C1). Participants' actual reporting was, however, far from this. Overreporting of high scores and underreporting of low scores was common. The highest pip score was reported almost four times as often as the lowest score ($M_{\text{max. pip}} = 30\%$ vs. $M_{\text{min. pip}} = 7\%$). By contrast, medium scores were reported about as frequently as would be expected from honest reporting. This does not mean, however, that participants who obtained medium scores were necessarily more honest than the rest. The total mean across all observations was largely consistent with “justified dishonesty” theory (Gächter & Schulz, 2016; Shalvi, Dana, et al., 2011), according to which participants respond as if they had rolled twice and then chose to report the larger outcome—although they were unambiguously instructed to report only the outcome of the first roll. According to justified dishonesty theory, even people who obtain medium scores will be willing to overreport if the second roll yields a greater score than the first.

Matrix Tasks

It has been suggested that participants in the matrix task inflate their results only to the extent that they can maintain a positive

(Appendices continue)

Table C2
Predictors of the Standardized Report (M_R) in Sender–Receiver Games: Regression Analyses With and Without the Number of Options as Covariate

Predictor (reference category)	M_r	M_r
Intercept	41% (5.27) [†]	37% (5.40) ^{***}
Investigative setting (laboratory)		
Online/telephone	–11% (6.05)	–12% (5.97) [*]
Field experiment	–14% (4.53) ^{**}	–14% (4.58) [*]
Participant characteristics (noneconomics students)		
Nonstudents	18% (5.55) ^{**}	16% (5.53) ^{**}
Economics students	–3% (3.95)	–3% (3.90)
Mechanical Turk workers	26% (6.89) ^{***}	27% (6.80) ^{***}
Normative cues	–20% (3.47) [†]	–20% (3.43) ^{***}
Experimental deception	–12% (3.44) ^{***}	–9% (3.62) [*]
Maximal externality	–1% (.45) [*]	–1% (.45) [*]
Maximal gain	1% (.45) ^{**}	1% (.45) ^{**}
Options	—	1% (.56) [*]
Observations	$k = 165$ $n = 9,243$	$k = 165$ $n = 9,243$
Residual heterogeneity	$I^2 = 84\%$ $\tau^2 = .02$	$I^2 = 84\%$ $\tau^2 = .02$
Heterogeneity accounted for	$R^2 = 37\%$	$R^2 = 39\%$

Note. Linear regression models with random effects at the experiment level. Unless denoted otherwise, values refer to beta weights with standard errors in parentheses.

[†] $p < .0001$. ^{*} $p < .05$. ^{**} $p < .01$. ^{***} $p < .001$.

self-concept of being an honest person in general (Mazar et al., 2008) or a positive social image (Abeler et al., 2016; Batson et al., 1997; Dana et al., 2007). That is, people shy away from reporting that they have solved all matrices (maximal lying) and instead inflate the number of solved matrices to some degree (truth stretching; see also Schweitzer & Hsee, 2002; Tenbrunsel & Messick, 2004). Our results are largely in line with this idea (Figure C2), although other explanations are possible (see General Discussion). The 48% of participants who reported dishonestly in matrix tasks for which we had the primary data on average claimed to solve “only” an additional 4 matrices ($SD = 4.65$) of an average of 14 unsolved matrices ($SD = 4.58$).

In the main text, the regression model predicting the standardized report in matrix tasks had a relatively poor model fit ($R^2 = 20\%$; Table 4), potentially due to the outcome measure itself. A better outcome measure than the standardized report could be the absolute number of unsolved matrices claimed as solved. To validate the regression model presented in Table 4, we therefore fitted an additional regression model to matrix tasks (Table C3) for which the dependent variable was the number of unsolved matrices claimed as solved. Moreover, we tested whether additional controls would change the results (the total number of matrices per task; total matrices; the percentage of matrices without a solution; percent unsolvable; and the time in minutes that participants were allotted for each matrix: time per matrix). Essentially, the results of all three models were qualitatively similar, with one exception: In

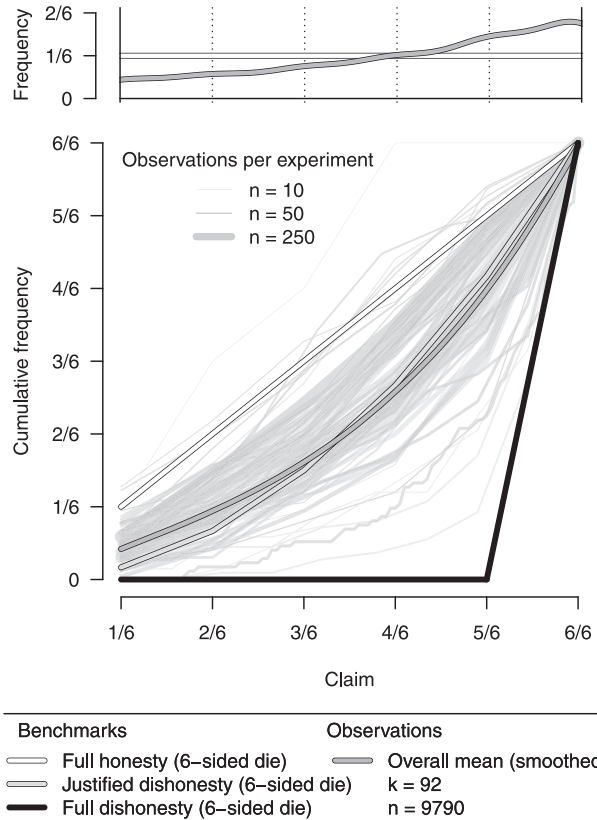
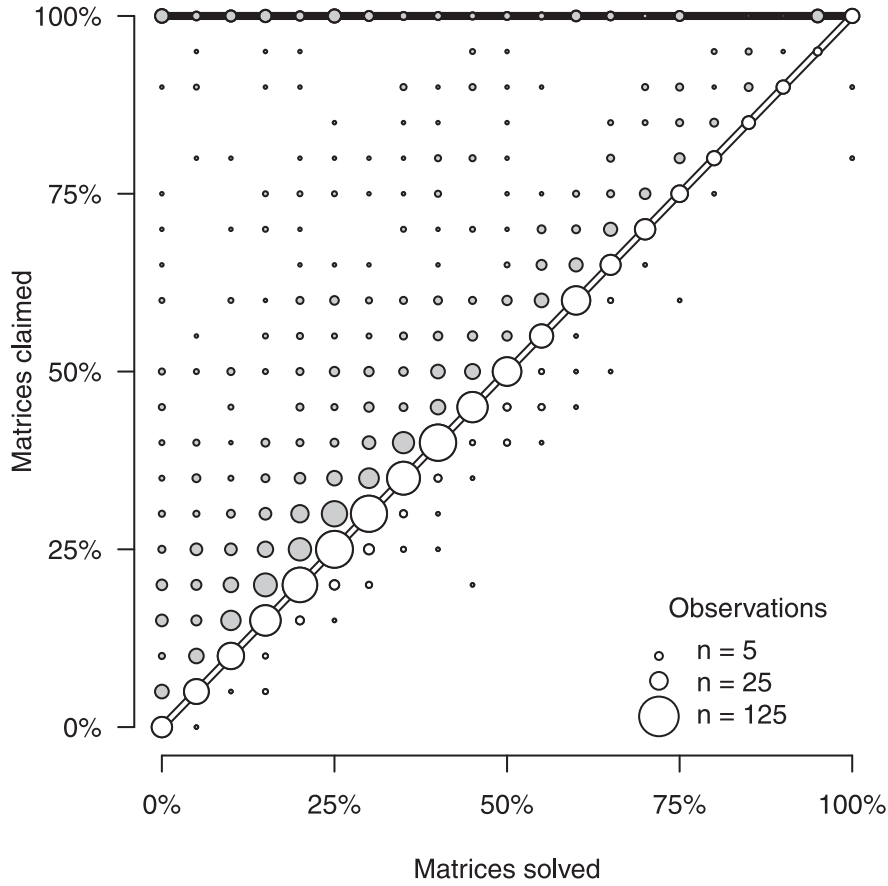
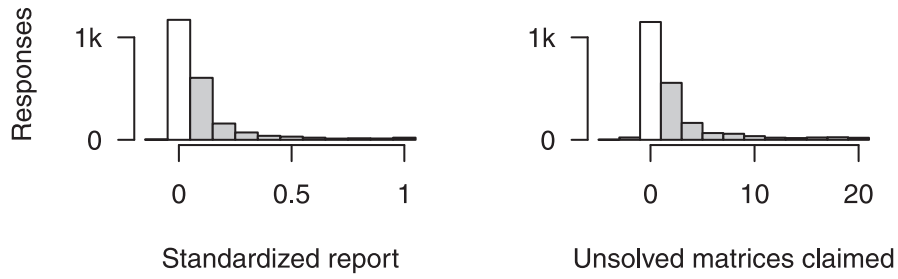


Figure C1. Distribution of reported scores in die-roll tasks with a single roll. The top graph shows the density function of the reported scores. The main graph shows the cumulative distribution functions (CDFs) of the reported score. The gray lines depict the CDFs per primary experiment. The more observations per experiment, the wider the line. The dark gray line represents the mean CDF for all experiments. The mean CDF was smoothed by local polynomial regression fitting and weighted by the number of observations per reported score. The full honesty benchmark (white) depicts the CDF for honest reporting. The justified dishonesty benchmark (light gray) represents the CDF for rolling twice and then reporting the greater score of the two rolls, instead of reporting the first score, as instructed. The full dishonesty benchmark (black) depicts the CDF for a sample of maximal liars.

contrast to the uncontrolled model predicting the standardized report, field experiment setting reached conventional levels of significance once controls were accounted for (from $p = .155$ to $p = .006$). Overall, the model fits improved only slightly with the additional controls. If anything, predicting the standardized report had a better fit than predicting the number of unsolved matrices claimed as solved. Note that the standardized report and the number of unsolved matrices claimed as solved were highly correlated (Spearman’s $\rho = 0.99$, $p < .001$), suggesting that the two measures provided essentially the same assessment of dishonesty.

(Appendices continue)



Benchmarks	Response type	Total observations
○ Full honesty	○ Dishonest report	k = 41
● Full dishonesty	○ Honest report	n = 2163

Figure C2. Number of claimed and solved matrices in the matrix task. The main plot depicts the proportion of matrices claimed as solved as a function of the proportion actually solved for experiments that measured performance and claims per participant and for which we had the primary data. Bigger dots stand for more observations. The top-left histogram depicts the distribution of standardized reports. The top-right histogram depicts the number of unsolved matrices claimed as solved—that is, the number of actually solved matrices subtracted from the number of matrices reported as solved.

(Appendices continue)

Table C3

Predictors of Different Outcome Measures in Matrix Tasks: Regression Analyses for Standardized Report (M_R) and Absolute Number of Unsolved Matrices Claimed as Solved

Dependent variable (reference category)	M_r	M_r	Unsolved matrices claimed
Intercept	25% (12.92)	128% (47.54)**	3.65 (1.80)*
Investigative setting (laboratory)			
Online/telephone	-13% (13.50)	-15% (13.54)	-1.72 (1.88)
Field experiment	-11% (7.77)	-30% (11.07)**	-1.88 (1.00)
Participant characteristics (noneconomics students)			
Nonstudents	-5% (13.16)	-5% (13.26)	-1.08 (1.09)
Economics students	5% (6.53)	-8% (8.21)	.41 (.86)
Mechanical Turk workers	—	—	—
Normative cues	-10% (5.94)	-11% (5.98)	-1.42 (.73)
Experimental deception	-3% (3.19)	-3% (3.17)	-.39 (.42)
Maximal externality	0% (.24)	-0% (.24)	.01 (.03)
Maximal gain	0% (.05)	-0% (.05)	-.01 (.01)
Additional controls			
Total matrices	—	-5% (2.04)*	—
Percent unsolvable	—	24% (12.43)	—
Time per matrix (in min)	—	-10% (47.71)	—
Observations	$k = 101$ $n = 6,093$	$k = 101$ $n = 6,093$	$k = 101$ $n = 6,093$
Residual heterogeneity	$f^2 = 77\%$ $\tau^2 = .01$	$f^2 = 76\%$ $\tau^2 = .01$	$f^2 = 77\%$ $\tau^2 = 2.28$
Heterogeneity accounted for	$R^2 = 20\%$	$R^2 = 22\%$	$R^2 = 16\%$

Note. Linear regression models with random effects at the experiment level. Unless denoted otherwise, values refer to beta weights with standard errors in parentheses.

* $p < .05$. ** $p < .01$.

Received September 20, 2017

Revision received September 10, 2018

Accepted September 12, 2018 ■