Pre-print
Working paper posted prior to peer review (June 2024).

Beliefs versus Reality: People Overestimate the Actual Dishonesty of Others

Jareef Martuza[1], Helge Thorbjørnsen[1], Hallgeir Sjåstad[1]

[1]Norwegian School of Economics, Department of Strategy and Management, Norway

Abstract

Beliefs about the dishonesty of other people can shape our collective behavior in powerful ways. How accurate are these beliefs? And do we believe that other people are similarly, more, or less dishonest than they truly are? In a research program on moral decision-making consisting of 31 different effects (N = 8,127), initially designed to test a broad collection of different hypotheses, participants were placed in various situations where they could lie for personal gain, without any repercussions or detection risk. Crucially, we also asked all participants to estimate what percentage of other people would lie in a similar situation. Conducting an internal meta-analysis across all experiments, including both incentivized choice experiments and hypothetical marketplace scenarios, the results revealed a substantial overestimation of others' dishonesty by an average of 14 percentage points (meta-analytic effect: Hedge's g = 0.58). That is, people are *less dishonest* than we tend to think, in which about 70% of our participants told the truth despite having a financial incentive to lie. These findings reveal a pervasive tendency to overestimate the actual dishonesty of other people, suggesting a biased belief that the world is less moral than it truly is. Does belief accuracy matter? In a new experiment using an information treatment based on our meta-analytic results (N = 981), providing correct information about actual honesty levels did not only lead to more positive honesty beliefs, but also to a broader enhancement of pro-social expectations – including higher interpersonal trust, fairness and helpfulness beliefs, and reduced cynicism.

Keywords: dishonesty; beliefs; meta-analysis; information treatment.

Open Practices Statement:
For the information experiment in the paper and four robustness experiments in the Online Supplement, the primary hypothesis and statistical analyses were formally pre-registered in advance of the data collection. Data, code, and study materials are openly available on the OSF platform (link: https://osf.io/dc63t/?view_only=9cd2b99d441543fa957a71a18c72f844).

Beliefs versus Reality:
People Overestimate the Actual Dishonesty of Others

In the current research, we study one simple question: Do people believe that others are similarly, more, or less dishonest than they truly are? Moral decision-making in general, and truth-telling in particular, is currently one of the most active research areas in behavioral social science. Ranging from psychology (Coltheart et al., 2011), economics (Schotter & Trevino, 2014), sociology (Kluegel & Smith, 1981), organizational behavior (Greenberg, 2002), and consumer behavior (Rotman et al., 2018), there is now an extensive literature studying the frequency of dishonest behavior, and what specific characteristics of the social context and individual person that are most important in shaping the basic willingness to cheat others for personal gain (Abeler, 2019; Gerlach et al., 2019). However, far less is known about the *beliefs* people have about the dishonesty of others.

In our view, beliefs about dishonesty are equally important to study as actual dishonesty, because most of our actions are filtered through our mental representations of the world, rather than the external reality directly. That is, it does not help much if most people are honest, if the average person does not *believe* that other people will tend to tell the truth. Conversely, if people should turn out to be less honest than we think, that would also be important to know, as it could create its own set of problems – including a social environment of exploitation, free-riding, and other violations of mutual trust.

The current paper will therefore do the following. First, we will conduct a complete internal meta-analysis of 11 recent experiments on moral decision-making, providing high-quality data that enables us to compare *actual dishonesty* with *beliefs about dishonesty* in the exact same test situation across 31 effect sizes from more than eight thousand participants. Second, we will conduct experimental robustness tests of the main finding from the meta-analysis, showing that the conclusion does not rely on potential order effects or specific wording of the belief measure (fully reported in the Online Supplement). Third, we will conduct a new experiment to examine whether informing people about the actual honesty levels of other people, based on the main results from our meta-analysis, will

have a causal effect on their social beliefs and pro-social expectations. If so, that would suggest that miscalibrated beliefs about dishonesty can have broad implications for how people think about the social world.

1. Actual Dishonesty

Honesty is like the air around us: We do not usually notice its presence, but its absence has detrimental consequences. Charles Ponzi's pyramid scheme, the Enron accounting scandal, and Sam Bankman-Fried defrauding crypto investors are just a few examples of severe dishonest behavior in modern times. While honesty is positively associated with trust, cooperation, and economic prosperity (Gächter & Schulz, 2016), dishonest behavior inflicts great costs on society. For example, merchandise returns fraud cost US retailers approximately $23.2 billion in 2021 (National Retail Federation, 2022), and about 13% of annual tax revenue ($428 billion) in the United States goes uncollected (Internal Revenue Service, 2022).

Beyond extreme fraudsters and career criminals, also "mostly honest" ordinary people may commit tax evasion (Bott et al., 2020), consumer fraud (National Retail Federation, 2022), or offer bribes (Sulitzeanu-Kenan et al., 2022). The dishonesty of everyday people has therefore increasingly captured the attention of behavioral scientists, both for theoretical and societal reasons. While Becker's (1968) theory of rational crime predicts that people will cheat as long as the material benefits exceed expected costs, it is now well-established in psychology and behavioral economics that people rarely cheat to the maximum extent (Abeler et al., 2019), essentially "leaving money on the table" even under anonymity and zero punishment (Gerlach et al., 2019; Rosenbaum et al., 2014).

From a social science perspective, an important question is therefore not only why people lie under some conditions, but also why self-serving dishonesty is not even more common than it is. Research on moral decision-making has examined the role of incentives for dishonesty (Wang & Murnighan, 2017; Wiltermuth, 2011), social context (Ayal et al., 2019; Cohn et al., 2022; Kocher et al., 2018), traits of the wrong-doer (Pascual-Ezama et al., 2015; Vincent et al., 2013), characteristics of the victim (Rotman et al., 2018; Soraperra et al., 2019), and the role of internal (self-image) versus

external (social image) motives in regulating dishonesty (Abeler et al., 2019). Another important stream of research has examined how to promote honesty using different behavioral interventions (Bicchieri et al., 2023; Bott et al., 2020), albeit with mixed findings and a lack of insights into underlying processes when something worked or did not work (Hertwig & Mazar, 2022; Kristal et al., 2020; Martuza et al., 2022; Skowronek, 2022).

Taken together, the study of behavioral dishonesty is prevalent and draws broad interest. From newsrooms and dinner tables to research labs and economic processes in society, it is a topic of conversation and academic discourse. Of particular interest for the current research, is to gain a better understanding of whether people's beliefs about dishonesty are generally accurate, or whether they might be systematically biased toward overestimation or underestimation of actual dishonesty.

2. Beliefs about Dishonesty

People tend to hold some fundamental beliefs about the social world, including how similar others are to the self (Robbins & Krueger, 2005), how favorable they are (Tarrant et al., 2012), and how to attribute the behaviors of others (Hewstone, 1990). In general, beliefs can shape our expectations and behavior in important ways (Jervis, 2006), such as the decision to cooperate or not in a public goods dilemma (Weber et al., 2023), using fare systems in public transport (Galeotti et al., 2021), or to rent an Airbnb apartment from a stranger based on their racial group identity (Nødtvedt et al., 2021). Importantly, beliefs can also affect behavior when they are wrong or poorly calibrated, meaning that they deviate from the best available evidence about social reality. As a timely example, during the beginning of the election year of 2024, American citizens have been *feeling* dissatisfied about the state of the economy, expressing strong beliefs about a spiraling negative trend, leading them to widespread disapproval of their political leaders and harboring a pessimistic outlook of the future. This is a striking pattern, because the best available empirical data suggests that the US economy has performed quite well in terms of GDP growth, stock market returns, and low unemployment during the same period (Pew Research Center, 2023). So, although beliefs and external reality are presumably positively correlated over time, in some cases they may differ drastically, in

which beliefs about the economy can be a stronger predictor of personal and democratic behavior than the actual economy (Dias et al., 2023).

Similarly, people also hold beliefs about how honest or dishonest other people tend to be (Bicchieri et al., 2023; Dimant & Gesche, 2023). For instance, beliefs about others' moral inclinations can shape strategic dishonesty in groups (Verrina et al., 2024), pre-emptive actions in curbing employee misconduct (List & Momeni, 2021), facilitating transactions among strangers (Köbis et al., 2021), and even influence the decision of asking strangers for help (Zhao & Epley, 2022). On the one hand, believing others to be less dishonest than they truly are and developing unwarranted trust can lead to naïve behaviors and policies vulnerable to exploitation by bad actors (Köbis et al., 2021). On the other hand, if people believe that others are more dishonest than they truly are, the public may develop unwarranted distrust, which may hinder economic growth (Algan & Cahuc, 2013), state functions (Herreros, 2023), and even social relations (Schilke et al., 2021). Thus, it is important to not only study behavioral dishonesty, but to also assess the realism of people's beliefs about dishonest behavior.

Research on the "moral forecasting error" compared moral beliefs with actual behavior (Teper et al., 2011), in a way that is directly relevant to the current work. In their experiment, individuals were randomly assigned to either make a prediction about how they would act in a moral dilemma, or given the opportunity to make a choice in that same situation. The results showed that participants in the behavior condition cheated significantly *less* for personal gain than participants in the prediction condition thought they would have done in the same situation. Although the study compared estimates of predicted moral behavior to actual behavior, the current research differs from Teper and colleagues' (2011) work as their study focused exclusively on potential forecasting errors in self-prediction. That is, how predictions about one's own moral behavior may differ from actual behavior, and not how moral beliefs about other people compare to their actual behavior. We therefore consider previous research on the moral forecasting error an important first step, but where the current paper will expand the scope by studying the accuracy and potential bias of *social beliefs*. Given the many differences in

judgments about oneself versus others (Cusimano & Goodwin, 2020; Molouki & Pronin, 2015; Vazire, 2018), we consider it to be an open question whether we will find overestimation or underestimation when aggregating all our available data on predictions about the dishonesty of other people, and compare that to people's actual behavior in the exact same situation.

Regarding moral beliefs about the behavior of others, there is some research from a negotiation context finding that participants overestimate the percentage of other participants who would deceive their negotiating counterparts to advance self-interest (Mason et al., 2018). However, Mason and colleagues (2018) studied these beliefs in a strategic situation where their own outcomes (gains/losses from the negotiation) were directly affected by the behaviors of those the beliefs are held about (negotiation counterpart), which provides less clear evidence on the pure accuracy of social beliefs. Put simply, pessimistic beliefs about others in a negotiation context can arise from self-preservation motives. In contrast, our paper focuses on general beliefs about others' dishonesty across various contexts, where personal outcomes are not influenced by those whom the beliefs are about.

Recent research also shows that people wrongly believe that society as a whole has become less moral over time (Mastroianni & Gilbert, 2023). Although people's reports of the current morality of their contemporaries have been stable across a 55-year span (1965-2020), at the same time, people tend to believe that their contemporaries have become less moral than they used to be -- an "illusion" of moral decline that has been robust across a 70-year span (1949-2019). Importantly, this particular study assessed beliefs about historical *change* in others' morality over time, compared to reports of the current actual morality of their contemporaries in the same period, without directly comparing moral beliefs about the absolute level of specific behaviors to actual behavior in the same situation. As such, the extensive work by Mastroianni and Gilbert (2023) provides relevant background for how people tend to mistakenly perceive historical trends, seeing a decline even when there is none, but it cannot empirically assess whether people have accurate beliefs about absolute honesty levels in the here and now. In the current research, we will compute a meta-analytic quantitative estimate of actual

behavioral dishonesty across several specific tasks, and then compare that proportion of actual dishonesty with people's beliefs about dishonesty in the exact same situation.

3. Beliefs vs. Reality: Competing Hypotheses

In our view, it is currently an open question whether people believe that others are similarly, more, or less dishonest than they truly are. Indeed, there is both empirical and theoretical basis to form competing hypotheses.

For instance, a global study on civic honesty found that people were more likely to return a lost wallet when the wallet contained more money than less money or no money, suggesting that the increase in moral motivation to return it was stronger than the increasing financial self-interest for higher money amounts (Cohn et al., 2019). In contrast to this behavioral pattern, neither academic experts (economists) nor non-experts (general population) were able to predict this result, in which non-experts believed that the exact opposite thing would happen (i.e., fewer returned wallets for higher money amounts). These findings suggest that both lay people and social scientists may have inaccurate beliefs about the moral behavior of other people, and that a more comprehensive empirical approach is needed to assess the accuracy or potential direction of a bias in moral beliefs.

Unlike the current research, the lost wallet study was conducted in the field, and the participants making predictions did not see or directly experience the same test situation themselves. Thus, there might have been several unobserved differences between the prediction scenario and the actual behavior scenario operating as potential confounds or at least competing explanations for the observed difference between predictions and behavior. To avoid that problem, the current meta-analysis will only focus on controlled incentivized experiments and survey experiments in which all participants who are making a prediction about the moral behavior of other people have just been part of the exact same test situation themselves.

The approach in the current paper provides a unique comparison of moral beliefs to actual behavior, combining 31 different effects across more than eight thousand study participants, keeping

all else constant. As part of an extensive research program in our lab group about moral decision-making, these experiments were initially designed to test a broad range of different hypotheses belonging to different projects, focusing on the role of contextual factors in shaping dishonesty. By focusing on a complete set of our own primary and secondary data on beliefs about others' dishonesty and the actual dishonesty level observed in each study, we have therefore created a unique data set by including and collating all direct experimental comparisons in one meta-analysis.

Moving to the specific propositions in our meta-analysis, the null hypothesis is that there will be no significant difference between moral beliefs and actual behavior in the domain of dishonesty. If true, that would presumably be "good news" for citizens, leaders, and policymakers, as it would suggest that the beliefs we have about each other are accurately calibrated with empirical reality – providing a solid basis to make well-informed decisions on topics of trust, cooperation, and control.

In contrast with that view, a set of competing hypotheses suggests that beliefs and behavior will systematically deviate. According to what we call the underestimation hypothesis (H1), people will underestimate actual dishonesty by predicting that other people are less dishonest than they truly are. This finding would be consistent with research from other life domains on optimism (Sharot et al., 2011) and wishful thinking (Babad, 1987; Babad & Katz, 1991), in which people tend to form rosy beliefs about other people and the future that systematically exceeds empirical reality, including "defaulting" to the truth when judging information from others (Levine, 2014, 2022).

According to what we call the overestimation hypothesis (H2), the exact opposite will happen: On average, people will predict that other people are more dishonest than they actually are. If true, this finding would be consistent with the moral forecasting error in self-prediction (Teper et al., 2011), and recent work on the illusion of moral decline concerning perceived behavioral change over time (Mastroianni & Gilbert, 2023).

4. The Current Research

A study seeking to provide a careful examination of the accuracy or potential bias in moral beliefs, should measure both the beliefs about dishonesty and actual dishonesty rates as similarly as possible. Ideally, this should be studied in a controlled test environment where a mismatch between beliefs and actual dishonesty does not stem from other differences across the prediction and behavioral contexts. Second, potentially confounding factors, such as fear of detection, punishment, and reputation should also be excluded, as such test situations can lead people to estimate how concerned others would be about external factors rather than estimating people's internal moral motivation to tell the truth. Third, research comparing beliefs about dishonesty with actual dishonesty should test the robustness of this comparison across different study contexts. This is important because of the wide range of dishonest behaviors people can perform, which makes it an additional contribution to examine the generalizability of miscalibrated beliefs about others' dishonesty, rather than relying on one specific context like returning wallets (Cohn et al., 2019), negotiations (Mason et al., 2018), or overreporting incentivized performance on a math task (Teper et al., 2011) as the only measure of dishonesty.

In trying to overcome these challenges, the current paper presents an exhaustive internal meta-analysis ($k = 31$, $N = 8{,}127$) of our effect of interest, comparing beliefs about others' dishonest behavior with actual rates of dishonest behavior in the exact same situation. These effects come from *all* experiments conducted as part of a research program on moral decision-making spanning three years (2022-24), which were initially designed to test a broad collection of different hypotheses about the drivers and mechanisms of behavioral dishonesty. This approach provides complete access to all relevant data where average beliefs about others' dishonesty and rates of actual dishonesty could be compared. In addition to the inclusion of effects from a range of contexts of dishonest behavior (e.g., the mind game, scenario decisions of returns fraud, or underbilling), the current meta-analytical approach is also free from the validity threat arising from the systematic omission of null results in the published literature (Sterling, 1959; Sterling et al., 1995) and unobservable "researcher degrees of freedom" (e.g., excluding observations, transforming measures, adding control variables) when conducting analyses across study (Simmons et al., 2011). Further, the internal validity of our meta-

analysis is bolstered by the fact that we included all relevant studies and conducted exactly one type of analysis across all of them, comparing beliefs about others' dishonest behavior to rates of dishonest behavior (for a similar meta-analytic approach from other research domains, see e.g., Pennycook & Rand, 2022; Vosgerau et al., 2019).

Based on the results of our meta-analysis, we will then ask a second research question in this paper: Does the accuracy or potential bias in honesty beliefs make any difference in how people think about the social world? To answer this question, we conducted a new experiment as our second main study in this paper, testing whether providing people with correct research-based information about the main results from our meta-analysis (actual honesty levels and the accuracy in people's beliefs about others' dishonesty) would influence their social beliefs more generally, including interpersonal trust, fairness, helpfulness, reciprocity, and cynicism.

In the sections that follow, we will first report our meta-analysis comparing beliefs about dishonesty with actual dishonesty. Following that, we will report an information experiment where we describe our main results from this meta-analysis in simple words, testing the potential causal effect of receiving correct information about actual dishonesty levels on a broad set of pro-social expectations.

3. Main Study 1: Meta-Analysis

3.1. Method

We will now present our internal meta-analysis comparing moral beliefs about the frequency of dishonest behavior among other people to the corresponding rate of actual dishonest behavior in the same participant samples. There were four criteria for an effect to be included in this meta-analysis: (1) The study was conducted by at least one of the three authors of this paper for a project part of a research program on moral decision-making (e.g., behavioral dishonesty against big vs. small business organizations), (2) The study contained a general measure about beliefs about the rate of dishonesty of other participants completing the survey (e.g., "What percentage of participants do you think, completing the same study as you, would falsely state (intentionally) that they guessed correctly?"), (3) Data from the same study could estimate the rate of actual dishonest behavior (e.g.,

percentage of participants falsely reporting a correct guess to win a bonus), and (4) the smallest sample size per condition in the study was equal to or greater than 100 participants. This led to the inclusion of all conditions from 11 separate experiments conducted between 2022 and 2024, comprising 31 unique effects. Please see Table 1 for a brief description of the effects. All experiments from which effects were extracted were conducted by recruiting U.S. American participants from *Prolific*, a reputable online participant recruitment platform that generates high-quality data from fairly representative general population samples (Peer et al., 2021).

Turning to how the data for the meta-analysis was prepared, we will now briefly describe the procedure. For each effect, we computed the mean and standard deviation of participants' beliefs of what percentage of others, from 0 to 100[1], they thought would behave dishonestly. Then, we conducted 31 one-sample t-tests to extract each separate effect by testing the mean reported beliefs about others against the target value of the in-sample proportion of actual dishonest participants. Based on that, we calculated Cohen's d and subsequently Hedge's g with corresponding standard errors for each effect.

Of the 11 separate experiments, 9 measured actual dishonest behavior using adaptations of the mind game paradigm (Jiang, 2013) and the color game paradigm (Albertazzi, 2022). In the mind game paradigm (Jiang, 2013), participants guess in their minds the outcome of a die roll, and then self-report after the die roll whether they were able to guess the correct outcome. Correct guessing is incentivized with a bonus, which means that the majority of participants, who did not make the correct guess, have to make a trade-off between the moral motive to tell the truth and the economic incentive to lie. This approach allows for the estimation of group-level dishonest behavior, by subtracting the statistically expected proportion of correct guesses (e.g., 1/6 chance to guess the outcome of a six-sided die) from the proportion of reported correct guesses, and also enables

[1] Beliefs regarding effect sizes 28 and 29 were reported from 0 to 83 as we explicitly mentioned that 17% of participants will be statistically able to attain additional bonuses by reporting a correct guess of a future die roll *without* having to lie.

examination of people's *beliefs* about others' dishonesty in the same situation, without any form of detection risk, punishment, or reputation concerns. In the color game paradigm, participants are shown a random list of 60 color names on a screen that disappears after five seconds (making it unrealistic for anyone to read all colors), asked to select a color to have in mind, and then shown three new random lists containing four colors each, with each list being associated with a specific payoff. The colors displayed in the original 60 are never present in the new list, so any choice other than "Not in the lists" can be classified as a payoff-increasing lie, which allows for direction estimation of individual-level dishonest behavior.

Two experiments (May 2022 and July 2022) had a repeated measures design, such that participants were asked about their beliefs about others on two rounds of an honest task. Nonetheless, we considered each belief measure across rounds and targets separately. In the two non-behavioral experiments (November 2022 and July 2023), we asked participants whether they intended to behave honestly or dishonestly in marketplace scenarios (merchandise returns fraud and not reporting being underbilled). Then, we asked them to estimate the percentage of others who would intend to behave dishonestly in these scenarios. This enabled us to compare how accurate people's beliefs are about others' moral intentions in the marketplace. Here too, we considered each belief measure regarding the rate of dishonest behavior against each target separately, and computed the beliefs-behavior difference measure by subtracting the percentage of dishonest choices from the participants' beliefs about the percentage of dishonest choices.

Table 1. Description of the experiments included in the meta-analysis.

Effect ID	Experiment Date	Cell size	What the study was originally designed to examine	Dishonesty measure	Dishonesty belief measure prompt	Sample Mean age (SD)	Belief minus reality Mean (SD)
1	May 2022	195	Differences in dishonest behavior against a big	Mind Game	What percentage of participants do you think would state that	27.4 (4.7)	10.7 (23.6)
2		190				27.5 (4.9)	17.3 (25.8)

3		195	vs. small business organization.		they guessed correctly?	27.4 (4.7)	14.0 (27.3)
4		190				27.5 (4.9)	20.1 (25.0)
5	July 2022	199	Differences in dishonest behavior against a participant with a male vs. female avatar	Color Game	As a potential bonus, what percentage of participants do you think will get A. £5 + £2 B. £5 + £5	35.1 (13.1)	14.0 (24.3)
6		199				35.1 (13.1)	14.2 (25.9)
7		202				35.2 (13.4)	9.3 (25.3)
8		202				35.2 (13.4)	25.3 (24.3)
9	August 2022	300	Differences in dishonest behavior against a big vs. small business organization	Mind Game	What percentage of participants do you think would state that they guessed correctly?	27.5 (4.8)	12.9 (22.3)
10		288				27.7 (4.7)	10.1 (22.3)
11	August 2022	308		Mind Game	What percentage of participants do you think would state that they guessed correctly?	38.5 (13.9)	8.6 (21.7)
12		283				38.3 (13.8)	6.1 (20.6)
13	November 2022	409	Differences in intention to commit "returns fraud" when paired with different types of chatbot avatars	Returns Fraud	Out of 100 participants, how many would you guess will choose a free return option with X? *10 participants with the closest guesses will get a £2 bonus.*	41.6 (14.6)	16.5 (21.6)
14		418				39.1 (13.8)	11.6 (21.6)
15		414				39.7 (14.0)	20.3 (23.3)
16		364				39.6 (13.6)	12.9 (21.4)
17	July 2023	291	Differences in intention to commit self-checkout fraud against big vs. small	Self-checkout	Imagine that 100 participants are taking this survey and are faced with the same decision. How many of them do you think	38.8 (14.7)	9.9 (21.1)
18		297				38.0 (14.7)	-3.0 (21.1)
19		294				36.1 (13.1)	-0.5 (21.6)

20		296	vs. medium vs. no-size signaling business		would choose to leave X Mart without reporting the billing error?	38.1 (14.4)	-1.0 (21.6)
21	March 2023	299	Differences in beliefs about others' dishonest behavior when asked before vs. after completing a dishonesty task	Mind Game	Out of 100 participants completing this task, how many do you think will report they guessed the same score as the die roll? Five participants with the closest guesses will receive an additional £5 within a week.	38.0 (13.0)	21.4 (23.3)
22		296				36.3 (12.2)	17.6 (25.1)
23	April 2023	205	Whether informed about erroneous beliefs about others' dishonest behavior, in various ways, influences subsequent dishonest behavior	Mind Game	We asked several American participants on Prolific to participate in a task last week…… What percentage of participants do you think reported they had guessed the same score as the die roll? Please enter a number only from 0 to 100, no text. Five participants with the closest estimates will receive an additional £5 within a week.	34.4 (11.5)	21.6 (23.7)
24		203				36.7 (12.9)	20.4 (26.2)
25		207				35.2 (11.8)	14.2 (25.6)
26	April 2023	301	Whether informed about erroneous beliefs about others'	Mind Game	We asked several American participants on Prolific to participate in a task earlier this month….	35.1 (12.1)	18.7 (25.8)
27		298				34.7 (12.2)	17.3 (26.5)

			dishonest behavior, in various ways, influences subsequent dishonest behavior		What percentage of participants do you think reported they had guessed the same score as the die roll? Please enter a number only from 0 to 100, no text. Five participants with the closest estimates will receive an additional £5 within a week.		
28	September 2023	292	Whether discrepancy between beliefs about and reality of dishonesty persist when explicitly mentioning others' dishonest behavior in the instructions	Mind Game	What percentage of participants do you think will falsely report they guessed the same score as the die roll? Please insert number only. Five participants with the closest estimates will receive an additional £5 within a week.	37.7 (12.)	26.2 (24.4)
29		291				38.1 (13.3)	24.2 (24.9)
30	Febraury 2024	100	Whether people are more likely to cheat their outgroup than their ingroup in artificial groups induced using a minimal group creation.	Mind game	What percentage of participants do you think, completing the same study as you, would **falsely state (intentionally) that they guessed correctly**? That is, among all participants who were <u>not</u> able to correctly guess	38.7 (12.9)	11.3 (22.2)
31		101				38.6 (12.6)	15.2 (25.3)

					whether it would be an odd or even number, what percentage of these participants do you think would falsely claim that they guessed correctly? Two participants with the closest guess will receive **$1** as additional bonuses within a week of completion of the survey.		

3.2. Results and discussion

We first estimate the overall effect size when comparing the potential difference in reported beliefs about others' dishonesty versus actual dishonest behavior. For each comparison, a one-sample t-test was conducted to compare the mean estimated percentage of dishonest others against the actual percentage of dishonest behavior, in which effect sizes were calculated as Hedge's g with corresponding standard errors. These standardized mean differences in dishonesty beliefs vs. behavior would comprise the input for the meta-analysis.

After completing compiling each effect size and relevant information in one datasheet in the appropriate format, we conducted a meta-analysis using the Meta-Analysis Module (MAJOR) in the open-source statistical software Jamovi (The jamovi project, 2022) and the "meta" package (Schwarzer et al., 2015) in R (R Core Team, 2022). We used each effect size in Hedge's g as the outcome, and the ID of the effect as the study label.

The results provided clear support for the *overestimation hypothesis*, in which people predict that other people are more dishonest than they actually are. Specifically, a random effects model (k = 31) showed that the meta-analytic estimate was positive and statistically significant (b = .58, SE = .05,

Z = 10.9, p < .001, 95% CI [.48, .68]. This suggests that participants overestimated what percentage of others were dishonest by 58% of a standard deviation, resembling an overestimation by 13.6 percentage points (weighted average). In other words, people significantly overestimated the percentage of dishonest others, reflecting a rather strong effect size.

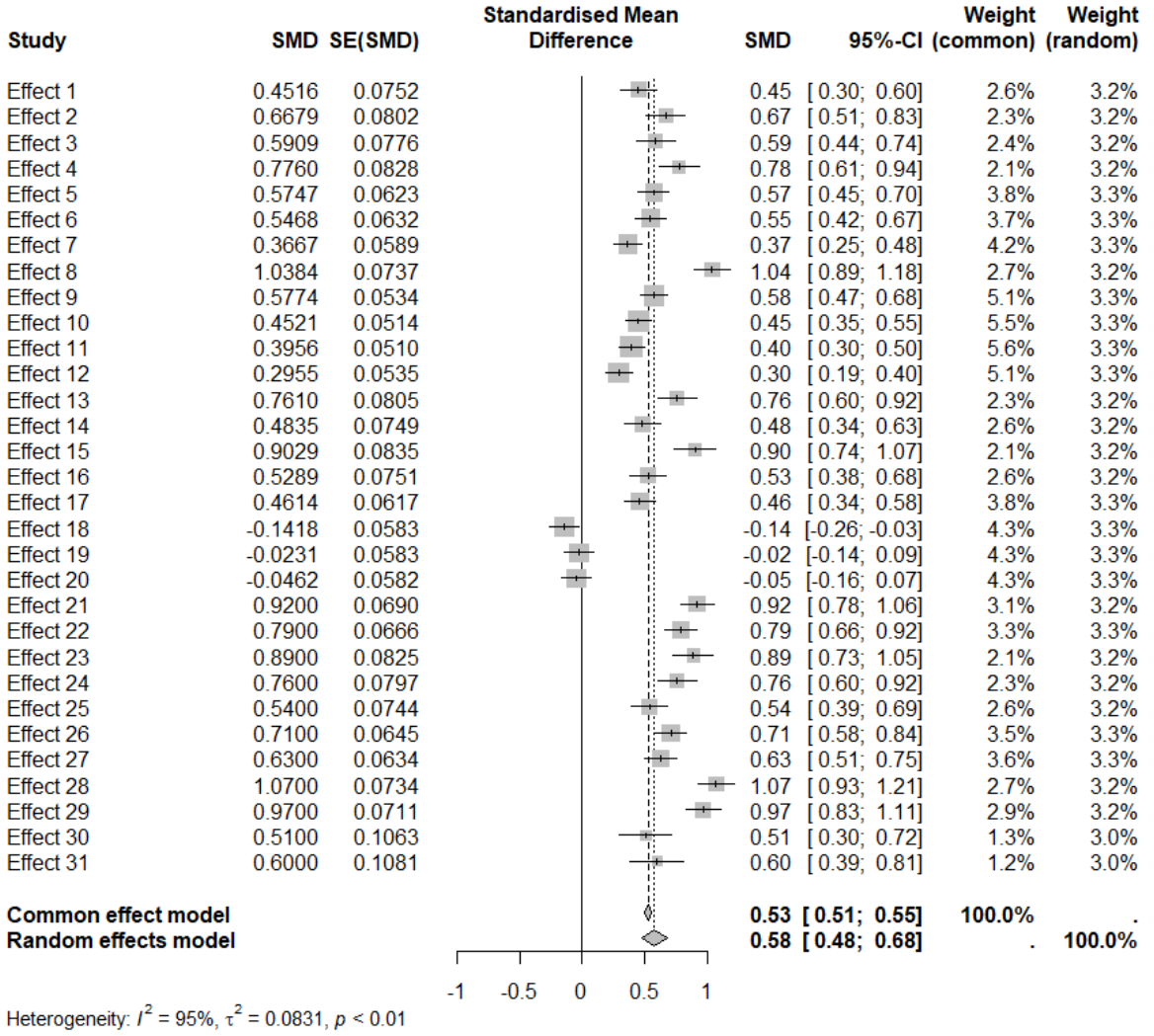

Figure 1. Meta-analysis (k = 31, N = 8, 127). The forest plot shows that across all comparisons, across 31 effects from 11 separate experiments, people significantly overestimated the actual percentage of dishonest others by 58% of a standard deviation (b = .58, SE = .05, Z = 10.9, p < .001, 95% CI [.48, .68]), which corresponds to a weighted average difference of 13.6 percentage points between predicted and actual rates of dishonesty.

We then conducted a heterogeneity analysis, finding significant heterogeneity in effect sizes across studies, Cochran's Q (30) = 626.9, I^2 = 94.8%, p < .001. Based on this result, we conducted a

series of separate moderator analyses[2] to gain a better understanding of this variation in effect sizes. First, we speculated that people's beliefs about marketplace dishonesty may be closer to reality than those about dishonesty in economic games, as people may be more likely to have prior information about others committing returns or self-checkout frauds (naturally occurring in the real world), compared to more abstract decision-making paradigms from experimental economics.

Accordingly, with the addition of a moderator regarding what context beliefs about others were reported (0 = economic game, 1 = marketplace behavior), a random-effect meta-analysis showed that the estimated intercept of the meta-analytic effect (b = .70, SE = .05, Z = 15.1, p < .001, 95% CI [.61, .79]) was qualified by a negative and statistically significant moderator (b = -.45, SE = .09, Z = -5.1, p < .001, 95% CI [-.63, -.28]. This suggests that although people overestimate the percentage of dishonest others overall, for decision contexts where participants may have less experience (economic games: stronger effect) people overestimate the percentage of dishonest others more than for contexts where participants may have more experience (marketplace behavior: weaker effect).

In addition, we also conducted two separate moderated meta-analyses to check for robustness across individuals and studies. First, with the addition of a moderator regarding whether dishonest behavior was detectable at the individual level (0 = no, 1 = yes), a random-effect meta-analysis showed that the estimated intercept of the meta-analytic effect (b = .71, SE = .06, Z = 12.1, p < .001, 95% CI [.59, .82]) was qualified by a negative and statistically significant moderator (b = -.32, SE = .09, Z = -3.5, p < .001, 95% CI [-.51, -.14]. This suggests that the overestimation of dishonest others decreased for decision contexts where dishonest behavior in principle could be detected (e.g., the color game rather than the mind game).

As a final potential moderator, we examined the possible role of whether the study was initially conducted to specifically examine the difference between beliefs and behavior (Effect IDs 21-29), meaning that the study was conducted specifically for this paper, (0 = no, 1 = yes). A random-

[2] We did not estimate all moderators together as their levels were not orthogonal.

effect meta-analysis showed that the estimated intercept of the meta-analytic effect (b = .49, SE = .06, Z = 8.7, p < .001, 95% CI [.38, .59]) was qualified by a positive and statistically significant moderator (b = .32, SE = .10, Z = 3.1, p = .002, 95% CI [.12, .53]. This suggests that the overestimation of the percentage of dishonest others increased for decision contexts in experiments conducted specifically for the current research. Nonetheless, the main effect was significant, bigger in magnitude, and in the same direction as the interaction. This means that although the overestimation of others' dishonesty was higher in experiments specifically conducted for the current paper, it was robustly significant in other experiments too that were initially designed to test different hypotheses.

Considering the meta-analysis of the main effects and moderators as a whole, we find that the overestimation of others' dishonesty is significant and robust on average, and does not get attenuated across different moderators. Although the strength of the main effect varied as a function of whether the study included familiar or new environments, potential detection risk, and was initially conducted for the current paper or not, the central finding was that the main effect of overestimating the actual dishonesty of other people was robust across studies. Next, we will briefly address two potential robustness issues by summarizing four follow-up experiments that are fully reported in the Online Supplement.

3.3. Robustness checks

To address potential methodological confounds in our meta-analytic estimate of dishonesty beliefs vs. behavior, we conducted four pre-registered experiments as follow-ups (Supplementary Experiments S1, S2a-b[3], and S3; methods and results are detailed in the Online Supplement[4]).

[3] All conditions in Experiments S2a-b were "belief" first conditions, and participants estimated the percentage of dishonest others having completed Experiment S1. The overestimation was robust, with average differences between beliefs and actual dishonesty being quite similar to that in Experiment S1.

[4] The effect sizes extracted from supplemental experiments (IDs 21-29) were included in the meta-analysis for full inclusion of all studies conducted in the research program. Nonetheless, the meta-analytic effect was robust when excluding effects from these experiments, showing these did not drive the average effect by themselves.

As a first potential confound, in most of the studies included in the meta-analysis (Effect ID's 1-20 & 30-31), participants' prior decisions could have influenced their subsequently reported beliefs about others. For instance, those who make a dishonest decision may overestimate others' tendency to do the same (being dishonest), as a form of rationalization or projection bias. Because the belief measure always followed the behavioral measure in these studies, we could not rule out prior decision effects on participants' beliefs about others. In Experiment S1 (pre-registration: https://aspredicted.org/VZN_PPB) in the Online Supplement, we used a between-subjects design to assess the belief-decision order empirically, by including a "belief first" condition where participants reported their beliefs about others before completing the task themselves, without being informed that they would have a chance to participate in the same task afterward. Replicating the main finding from the meta-analysis, the overestimation of others' dishonesty was found again, also when testing average beliefs in the "belief first" condition against the percentage behavior in the "decision first" condition. The average difference between beliefs and actual dishonesty corresponded to about 17.5 percentage points ($p < .001$).

As a second potential confound, in several studies included in the meta-analysis (Effect IDs 1-12, and 21-27), we did not explicitly ask the participants about the *dishonest* behavior of others. Instead, we indirectly measured beliefs about others' dishonesty by asking, for instance, what percentage of other participants would report a correct guess, rather than directly asking what percentage of other participants would falsely report a correct guess. To make sure the main finding in our meta-analysis does not reflect a comprehension problem in the way the belief question was asked, in Experiment S3 (pre-registration: https://aspredicted.org/KT5_RZB) in the Online Supplement, using a similar "decision first" vs. "belief first" between-participants design as Experiment S1, we told participants that the percentage of correct guesses, if reported truthfully, should be 17%, and then directly asked them "What percentage of participants do you think will falsely report they guessed the same score as the die roll? Please insert number only." Replicating the main finding in the meta-

analysis, participants overestimated the proportion of dishonest others by 24.2 percentage points[5], even when made explicitly aware of the probabilities regarding false reporting and directly asked about the percentage of dishonestly behaving others.

Together, Experiments S1 and S3 show that the overestimation of others' dishonesty is robust across both the order of belief elicitation vs. decision-making, and different wording of the instructions used to measure beliefs. In addition, we conducted two more robustness experiments focusing on asking people to estimate the proportion of dishonest others completing the study one week (Experiment S2a) or one month (Experiment S2b), in which we found similar results of significant and substantial overestimation (fully reported in the Online Supplement).

4. Main Study 2: Information experiment on pro-social expectations

In our internal meta-analysis, we found robust evidence for a tendency to overestimate the dishonesty of other people, identifying a systematic discrepancy between beliefs and reality. Do such beliefs matter for how we think about other people in general? There has been documented a wide range of positive correlates of honesty beliefs, including social outcomes such as trust (European Social Survey European Research Infrastructure (ESS ERIC), 2023) and cooperation (Fischbacher et al., 2001), but it remains an open question whether there might be a causal effect of that kind.

To assess this question empirically, we conducted our main Study 2 (pre-registration: https://aspredicted.org/6ZC_8FF) to examine whether informing individuals about the tendency to overestimate others' dishonesty, which is the main finding of our meta-analysis, might have a positive effect on their broader pro-social expectations to other people. If so, such a finding would suggest that biased and "uncorrected" dishonesty beliefs might be an obstacle to successful pro-social interaction, by making people more skeptical and distrusting of other people than what is warranted by the data.

[5] Comparing average beliefs in the "belief first" condition against estimated dishonest behavior in the "decision first" condition

In a between-subjects design, participants were randomly assigned to either a control condition or an information treatment condition. In this experiment, we tested the general hypothesis that exposure to research-based information describing the actual dishonesty level in our research would lead to enhanced pro-social expectations in general, reflected in higher beliefs in people's honesty, trustworthiness, helpfulness, and reciprocity, and reduced cynicism beliefs.

4.1. Method

1001 participants based in the United States were recruited using Prolific to participate in the study for a participation fee. The recruitment post targeted participants who were fluent in English, had at least a 95% approval rate on completing more than 100 tasks, and had an approval rate of at least 98%. After removing those who failed a simple attention check at the beginning of the study (N = 20), the final dataset had responses from 981 participants (Mean age = 40.2, SD = 13.0; 486 Male, 481 Female, 14 Other). A sensitivity power analysis using G*POWER (Faul et al., 2007) showed that the achieved sample size had 80% power to detect an effect size of d = .18 or larger in an independent samples t-test ($p < .05$, two-tailed).

At the start of the study, participants completed a simple attention check. Participants were presented with general instructions about the importance of reading instructions carefully. To show that they were attentive when responding to a general study instruction, participants were asked to select "A lot" to the question "How much do you like sports?", being when presented with five different options as potential answers.

Then, participants in both conditions were presented with information about how we as researchers have conducted studies on honesty with thousands of American participants. We also gave them an example of a die roll study where participants could lie to increase their payments without any possibility of being detected. After participants confirmed they had read and understood the general description of this research, they responded to a simple question of whether they thought one could learn something important from this type of research, before they could proceed.

After that, participants who were randomly assigned to the information treatment condition were presented with a second screen with additional information about the *results* from this research. Specifically, the information treatment involved presenting participants with two main findings from our meta-analysis. First, information about general honesty levels was presented as follows: "We find that the vast majority of people decide to tell the truth, corresponding to an average of 70% of our participants." Second, information about people's tendency to overestimate others' dishonesty was presented as follows: "We find that they mistakenly overestimate the dishonesty of other people by an average of 14 percentage points." At the end of this information treatment, participants were asked whether they had read and understood the description of the research, consisting of two comprehension-check questions asking about the reported percentage of honesty in the text (correct answer: 70%), and the accuracy of people's beliefs about others' dishonesty (correct answer: people overestimated it). Participants in the control condition received no such information or questions about the specific results, only the general description of this type of research on the first screen.

Moving on to the outcome measures, all participants in both conditions then responded to the same set of social beliefs measures, reflecting different kinds of pro-social expectations from the existing literature, by rating their agreement with each item on a scale from 0 (Strongly disagree) to 10 (Strongly agree). These measures consisted of a manipulation check asking about specific honesty beliefs ("Generally speaking, most people tend to tell the truth."), followed by broader measures of different pro-social expectations concerning trust beliefs ("Generally speaking, most people can be trusted."; (Delhey et al., 2011), fairness ("Most people try to be fair and not take advantage of others if they get a chance."; OECD, 2017), helpfulness ("Most people try to be helpful and not just look out for themselves."; OECD, 2017), reciprocity ("To help somebody is the best policy to be certain that he or she will help you in the future."; Perugini et al., 2003) and cynicism ("Kind-hearted people usually suffer losses."; (Leung et al., 2010). The order of presentation of these items was randomized across participants.

Finally, all participants responded to demographic measures of age, gender, education, income, and political orientation. Please see Table S5 in the Online Supplement for a full list of measures.

4.2. Results and discussion

As pre-registered hypotheses, we predicted that exposure to the information treatment would lead to higher honesty beliefs (manipulation check), higher trust beliefs (H1), higher fairness beliefs (H2), higher helpfulness beliefs (H3), higher reciprocity beliefs (H4), and reduced cynicism beliefs (H5). Our manipulation check was positive and significant, in which participants reported higher average beliefs in other people's honesty in the information treatment ($M = 6.50$, $SD = 1.91$) than the control ($M = 5.35$, $SD = 1.91$) condition, $t(979) = 8.55$, $p < .001$, $d = .55$. Thus, the information treatment providing accurate research-based information about actual honesty levels made participants believe that other people are more honest than participants in the baseline control condition thought.

The next and primary question then, is whether the information treatment would have a broader effect on pro-social expectations. In support of this idea and our five specific pre-registered hypotheses (H1-5), a series of independent samples t-tests found that social beliefs about other people were significantly more positive in the information condition than the control condition across all five measures (trust: $M = 5.81$, $SD = 2.18$ vs. $M = 5.09$, $SD = 2.17$, $t(979) = 5.17$, $p < .001$, $d = .33$), (fairness: $M = 6.26$, $SD = 1.98$ vs. $M = 5.38$, $SD = 1.98$, $t(979) = 6.97$, $p < .001$, $d = .45$), (helpfulness: $M = 6.16$, $SD = 1.98$ vs. $M = 5.42$, $SD = 1.94$, $t(979) = 5.58$, $p < .001$, $d = .38$), (reciprocity: $M = 5.96$, $SD = 2.48$ vs. $M = 5.60$, $SD = 2.51$, $t(979) = -3.33$, $p = .024$, $d = .14$), including lower cynicism beliefs (cynicism: $M = 5.77$, $SD = 2.56$ vs. $M = 6.29$, $SD = 2.29$, $t(979) = -3.33$, $p < .001$, $d = -.21$).

Although our primary analysis reported above tested five different pre-registered hypotheses, one could argue that they all measure different dimensions of pro-social expectations more generally. Therefore, one could also argue that it would be appropriate to control for multiple comparisons, since

both the manipulation check and each of the five specific hypotheses could potentially support the broader claim that providing correct information about dishonest behavior has a positive effect on pro-social expectations in general, thus increasing the risk of making a type 1 error (i.e., accepting a false-positive finding). In response to this concern, we conducted a Bonferroni correction as a non-registered robustness check, employing the simplest and most conservative method to account for multiple comparisons. For the Bonferroni correction, the pre-registered alpha ($p < .05$) is divided by the number of comparisons, which in our case is 5. This adjustment requires the alpha for the t-tests to be $p < .01$ ($\alpha = 0.05 / 5 \approx 0.01$) to be considered statistically significant. Based on this criteria, these results show that the treatment effect of providing information was highly robust for trust ($p < .001$), fairness ($p < .001$), helpfulness ($p < .001$), and cynicism beliefs ($p < .001$), under the new alpha of $p < .01$. However, the effect on reciprocity beliefs was no longer statistically significant after correction ($p = 0.024$). In other words, the information effect was robust in the predicted direction on four out of five different measures of pro-social expectations. Please see Figure 2 for an illustration of all comparisons across conditions and measures, side-by-side.

Taken together, these findings provide novel causal evidence that a simple information treatment, that informs people about the actual dishonesty rate among others and people's tendency to overestimate that rate, can change their social beliefs about other people in a positive direction.

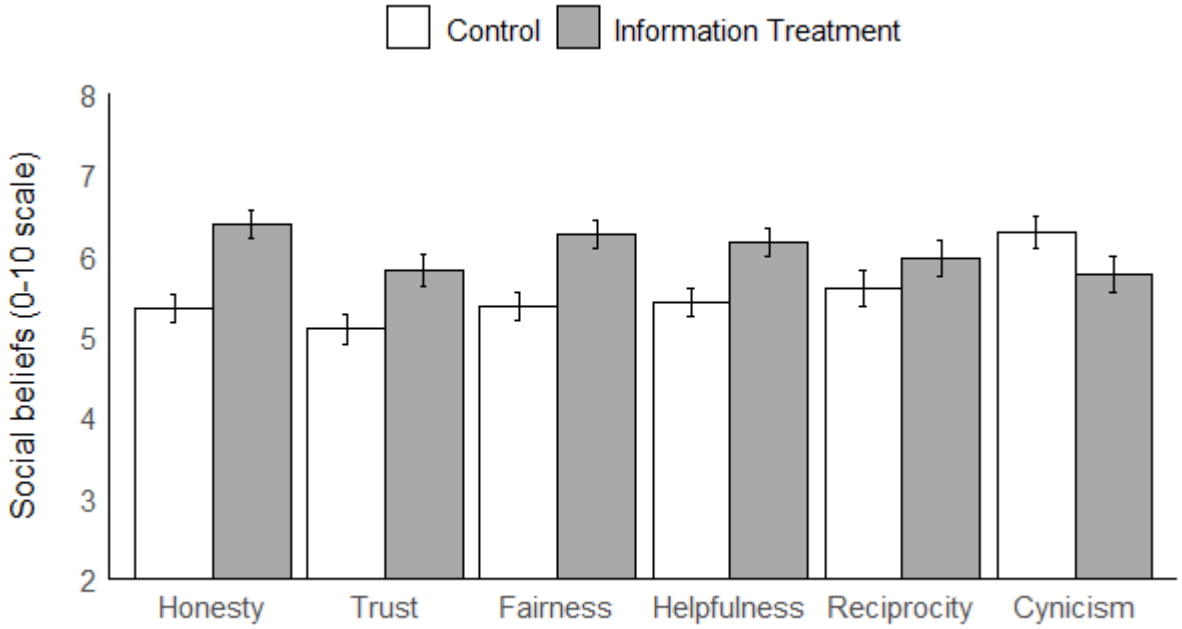

Figure 2. The effects of providing research-based information about actual honesty levels and the observed tendency to overestimate the dishonesty of others, Experiment 4 (N = 981). Compared to participants in the control condition, participants in the information treatment reported more positive honesty beliefs (manipulation check: $d = .55$, $p < .001$) and also reported more positive pro-social expectations (trust: $d = .33$, $p < .001$; fairness: $d = .45$, $p < .001$; helpfulness: $d = .38$, $p < .001$; reciprocity: $d = .14$, $p = .024$; cynicism: $d = -.21$, $p < .001$). When using an adjusted alpha level of 1% ($p < .01$) to correct for multiple comparisons, the information effect was robust and remained statistically significant on four out of five dimensions of pro-social expectations.

5. General Discussion

The current paper examines a fundamental question: Do people believe that others are similarly, more, or less dishonest than they actually are? Providing a high-powered empirical answer to that question, our meta-analysis revealed a significant *overestimation* of the actual dishonesty of other people by an average of 14 percentage points (meta-analytic effect: Hedge's $g = 0.58$), when comparing aggregate beliefs about dishonesty with actual dishonesty in the exact same test situation. This finding supports the overestimation hypothesis of dishonesty, stating that other people are *less* dishonest than we tend to think, and contradicts the underestimation hypothesis. In addition to a relatively large meta-analytic effect, based on 31 separate effect sizes and more than eight thousand

participants, our supplemental experiments showed that the overestimation effect persisted across belief-decision order and variations in instructions and measures, indicating that the main finding is robust to potentially important variations in methods and procedures.

Second, we wanted to know whether having biased beliefs about dishonesty plays a role in how we think about other people and the social world more generally. Does it really matter? In response to this question, we conducted a new experiment with about one thousand participants, and found that informing people about the actual honesty rate and the tendency to overestimate others' dishonesty led to enhanced social beliefs. Indeed, participants who learned about the main findings from our research program not only changed their views on dishonesty, but also expressed more positive pro-social expectations about other people in general – including higher interpersonal trust, higher fairness and helpfulness beliefs, and lower cynicism. This is important, as each of these beliefs has previously been linked to successful pro-social interaction in groups, organizations, and society.

The systematic overestimation of others' dishonesty can be integrated into the broader literature on social cognition and moral psychology, and has applied consequences for customer surveillance, employee monitoring, and interpersonal relations more generally. In short, when seeing our meta-analysis and the information experiment in combination, it seems likely that “uncorrected” negative beliefs about actual dishonesty levels in some cases can lead to missed opportunities for mutual cooperation, *unwarranted distrust*, and a tendency to prioritize top-down control over individual freedom. Moreover, our study challenges the 'wisdom of crowds' approach (Galton, 1907; Surowiecki, 2005; Van Dolder & Van Den Assem, 2017) to moral judgment, revealing that collective estimates regarding dishonest behavior can be biased toward pessimism. Our findings also contribute to understanding the mixed effects of social norm nudges, especially in tax evasion (John & Blume, 2018) and insurance fraud (Martuza et al., 2022): For individuals who are deeply entrenched in their biased beliefs about others’ dishonesty, this may lead them to disbelieve factually correct descriptive social norm messaging, such as “90% of your peers report correct information when filling forms”. To that end, a possible alternative in social norm nudging can be to provide research-based and

comprehensive evidence regarding the dishonesty of others, as we did in our information provision experiment.

In terms of applied implications and directions for future research, the systematic overestimation of others' dishonesty suggests that there is a need to develop and test trust-building interventions to correct misperceptions and biased beliefs. In our view, it is especially important that leaders and policymakers have access to an updated data-based view of the actual honesty levels in society, and do not rely too much on selective news headlines or their own gut feelings. Indeed, as we write this, businesses and society seem to be heading in the opposite direction, towards collective distrust: More and more supermarkets are putting even low-value products such as deodorant, toothpaste, and soap behind lock-and-key, and are generally increasing security measures (Meyersohn, 2022). In the workplace, a surge in monitoring employee activities may be fueling worker distrust (Christian, 2022), and may also have unintended side-effects by reducing the intrinsic worker motivation to do one's best (Falk & Kosfeld, 2006). Recognizing the tendency to overestimate others' dishonesty, firms may actually benefit from scaling back some security measures in the marketplace, and to assess the long-term effects empirically with an open and critical mind -- especially in self-service environments where surveillance costs can offset gains from reduced labor costs. Also within organizations, leaders may benefit from reconsidering employee policies based on mistrust, such as excessive monitoring during remote/hybrid work (Gartner, 2024).

Finally, we note that our findings in the current paper should be interpreted with several limitations in mind. First, the use of American participants from the United States in all our studies limits the generalizability across cultures, making it a question for future research whether the overestimation of actual dishonesty is equally high in different societies. Second, experimental honesty tasks in anonymous settings provide high precision in understanding the direct trade-off between moral motivation and financial self-interest, but may not fully capture the complexity and contextual richness of real-world dishonesty. Future research can therefore benefit from examining how estimates and actual rates of dishonesty compare in settings that mirror real-world decisions

more closely, for instance by correlating behavioral data from controlled experiments with the same from field settings (for a recent example, see: Dai et al., 2018). Third and finally, our reliance on mostly one-shot belief reports precludes analysis of within-individual variations across contexts and time. Future research could explore these dynamics further, in a study context of repeated interaction over time.

With these limitations in mind, we conclude that given the current study context and choice of research methods, we found robust meta-analytic evidence that people tend to overestimate the dishonesty of other people. In addition, we also found experimental evidence that correcting biased dishonesty beliefs can have real consequences for social thinking: Individuals who learned about the main findings from our research not only reported a stronger belief in the basic honesty of other people, but they also expressed more positive pro-social expectations in general – including higher trust, fairness and helpfulness beliefs, and reduced cynicism.

Acknowledgments

Our studies comply with all relevant ethical regulations regarding human research participants, including the guidelines from the Helsinki Declaration. As Norwegian laws and regulations do not require review by an institutional review board for anonymous, non-medical, low-risk research with human participants, this project was not submitted to such review. Informed consent was collected from all participants. We received partial funding for the current research from the Digital Innovation for Sustainable Growth (DIG) research center at the Norwegian School of Economics. The current research and manuscript significantly benefitted from the feedback from attendees of the SOL Wednesday seminar at NHH, the EMAC Doctoral Colloquium and Annual Conference 2024 (Bucharest), and the EGAGE EU Early-Stage Researcher Symposium at Luiss University.

References

Abeler, J., Nosenzo, D., & Raymond, C. (2019). Preferences for Truth-Telling. *Econometrica*, *87*(4), 1115–1153. https://doi.org/10.3982/ECTA14673

Albertazzi, A. (2022). Individual cheating in the lab: A new measure and external validity. *Theory and Decision*, *93*(1), 37–67. https://doi.org/10.1007/s11238-021-09841-0

Algan, Y., & Cahuc, P. (2013). Trust and Growth. *Annual Review of Economics*, *5*(1), 521–549. https://doi.org/10.1146/annurev-economics-081412-102108

Babad, E. (1987). Wishful thinking and objectivity among sports fans. *Social Behaviour*, *2*(4), 231–240.

Babad, E., & Katz, Y. (1991). Wishful Thinking—Against All Odds. *Journal of Applied Social Psychology*, *21*(23), 1921–1938. https://doi.org/10.1111/j.1559-1816.1991.tb00514.x

Becker, G. S. (1968). Crime and Punishment: An Economic Approach. In N. G. Fielding, A. Clarke, & R. Witt (Eds.), *The Economic Dimensions of Crime* (pp. 13–68). Palgrave Macmillan UK. https://doi.org/10.1007/978-1-349-62853-7_2

Bicchieri, C., Dimant, E., & Sonderegger, S. (2023). It's not a lie if you believe the norm does not apply: Conditional norm-following and belief distortion. *Games and Economic Behavior*, *138*, 321–354. https://doi.org/10.1016/j.geb.2023.01.005

Bott, K. M., Cappelen, A. W., Sørensen, E. Ø., & Tungodden, B. (2020). You've Got Mail: A Randomized Field Experiment on Tax Evasion. *Management Science*, *66*(7), 2801–2819. https://doi.org/10.1287/mnsc.2019.3390

Coltheart, M., Langdon, R., & McKay, R. (2011). Delusional Belief. *Annual Review of Psychology*, *62*(1), 271–298. https://doi.org/10.1146/annurev.psych.121208.131622

Cusimano, C., & Goodwin, G. P. (2020). People judge others to have more voluntary control over beliefs than they themselves do. *Journal of Personality and Social Psychology*, *119*(5), 999–1029. https://doi.org/10.1037/pspa0000198

Dai, Z., Galeotti, F., & Villeval, M. C. (2018). Cheating in the Lab Predicts Fraud in the Field: An Experiment in Public Transportation. *Management Science*, *64*(3), 1081–1100. https://doi.org/10.1287/mnsc.2016.2616

Delhey, J., Newton, K., & Welzel, C. (2011). How General Is Trust in "Most People"? Solving the Radius of Trust Problem. *American Sociological Review*, *76*(5), 786–807. https://doi.org/10.1177/0003122411420817

Dias, R., Sharma, E., & Fitzsimons, G. J. (2023). *Consumer Wealth and Price Expectations*. https://marketing.wharton.upenn.edu/wp-content/uploads/2023/09/10.19.2023-Dias-Rodrigo-PAPER-PriceExpectations-UnderReviewJCR.pdf

Dimant, E., & Gesche, T. (2023). Nudging enforcers: How norm perceptions and motives for lying shape sanctions. *PNAS Nexus*, *2*(7), pgad224. https://doi.org/10.1093/pnasnexus/pgad224

European Social Survey European Research Infrastructure (ESS ERIC). (2023). *ESS9—Integrated file, edition 3.2* [dataset]. Sikt - Norwegian Agency for Shared Services in Education and Research. https://doi.org/10.21338/ESS9E03_2

Falk, A., & Kosfeld, M. (2006). The Hidden Costs of Control. *American Economic Review*, *96*(5), 1611–1630. https://doi.org/10.1257/aer.96.5.1611

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*(2), 175–191. https://doi.org/10.3758/BF03193146

Fischbacher, U., Gächter, S., & Fehr, E. (2001). Are people conditionally cooperative? Evidence from a public goods experiment. *Economics Letters*, *71*(3), 397–404. https://doi.org/10.1016/S0165-1765(01)00394-9

Gächter, S., & Schulz, J. F. (2016). Intrinsic honesty and the prevalence of rule violations across societies. *Nature*, *531*(7595), 496–499. https://doi.org/10.1038/nature17160

Galeotti, F., Maggian, V., & Villeval, M. C. (2021). Fraud Deterrence Institutions Reduce Intrinsic Honesty. *The Economic Journal*, *131*(638), 2508–2528. https://doi.org/10.1093/ej/ueab018

Gartner. (2024). *Gartner HR Research Finds High-Performers, Women, Millennials Are Greatest Flight Risks When Strict Return to Office Mandates Are Implemented*. https://www.gartner.com/en/newsroom/press-releases/2024-01-30-gartner-hr-research-finds-high-performers-women-millennials-are-greatest-flight-risks

Gerlach, P., Teodorescu, K., & Hertwig, R. (2019). The truth about lies: A meta-analysis on dishonest behavior. *Psychological Bulletin*, *145*(1), 1–44. https://doi.org/10.1037/bul0000174

Greenberg, J. (2002). Who stole the money, and when? Individual and situational determinants of employee theft. *Organizational Behavior and Human Decision Processes*, *89*(1), 985–1003. https://doi.org/10.1016/S0749-5978(02)00039-0

Herreros, F. (2023). The State and Trust. *Annual Review of Political Science*, *26*(1), 117–134. https://doi.org/10.1146/annurev-polisci-051921-102842

Hertwig, R., & Mazar, N. (2022). Toward a taxonomy and review of honesty interventions. *Current Opinion in Psychology*, *47*, 101410. https://doi.org/10.1016/j.copsyc.2022.101410

Hewstone, M. (1990). The 'ultimate attribution error'? A review of the literature on intergroup causal attribution. *European Journal of Social Psychology*, *20*(4), 311–335. https://doi.org/10.1002/ejsp.2420200404

Internal Revenue Service. (2022). *Federal Tax Compliance Research: Tax Gap Estimates for Tax Years 2014-2016*. Research, Applied Analytics & Statistics. https://www.irs.gov/pub/irs-pdf/p1415.pdf

Jervis, R. (2006). Understanding Beliefs. *Political Psychology*, *27*(5), 641–663. https://doi.org/10.1111/j.1467-9221.2006.00527.x

Jiang, T. (2013). Cheating in mind games: The subtlety of rules matters. *Journal of Economic Behavior & Organization*, *93*, 328–336. https://doi.org/10.1016/j.jebo.2013.04.003

Kluegel, J. R., & Smith, E. R. (1981). Beliefs about Stratification. *Annual Review of Sociology*, *7*(1), 29–56. https://doi.org/10.1146/annurev.so.07.080181.000333

Köbis, N. C., Soraperra, I., & Shalvi, S. (2021). The Consequences of Participating in the Sharing Economy: A Transparency-Based Sharing Framework. *Journal of Management*, *47*(1), 317–343. https://doi.org/10.1177/0149206320967740

Kristal, A. S., Whillans, A. V., Bazerman, M. H., Gino, F., Shu, L. L., Mazar, N., & Ariely, D. (2020). Signing at the beginning versus at the end does not decrease dishonesty. *Proceedings of the National Academy of Sciences*, *117*(13), 7103–7107. https://doi.org/10.1073/pnas.1911695117

Leung, K., Ip, O. K. M., & Leung, K. (2010). Social Cynicism and Job Satisfaction: A Longitudinal Analysis. *Applied Psychology*, *59*(2), 318–338. https://doi.org/10.1111/j.1464-0597.2009.00392.x

Levine, T. R. (2014). Truth-Default Theory (TDT): A Theory of Human Deception and Deception Detection. *Journal of Language and Social Psychology*, *33*(4), 378–392. https://doi.org/10.1177/0261927X14535916

Levine, T. R. (2022). Truth-default theory and the psychology of lying and deception detection. *Current Opinion in Psychology*, *47*, 101380. https://doi.org/10.1016/j.copsyc.2022.101380

List, J. A., & Momeni, F. (2021). When Corporate Social Responsibility Backfires: Evidence from a Natural Field Experiment. *Management Science*, *67*(1), 8–21. https://doi.org/10.1287/mnsc.2019.3540

Martuza, J. B., Skard, S. R., Løvlie, L., & Thorbjørnsen, H. (2022). Do honesty-nudges really work? A large-scale field experiment in an insurance context. *Journal of Consumer Behaviour*, *21*(4), 927–951. https://doi.org/10.1002/cb.2049

Mason, M. F., Wiley, E. A., & Ames, D. R. (2018). From belief to deceit: How expectancies about others' ethics shape deception in negotiations. *Journal of Experimental Social Psychology*, *76*, 239–248. https://doi.org/10.1016/j.jesp.2018.02.013

Mastroianni, A. M., & Gilbert, D. T. (2023). The illusion of moral decline. *Nature*, *618*(7966), 782–789. https://doi.org/10.1038/s41586-023-06137-x

Molouki, S., & Pronin, E. (2015). Self and other. In M. Mikulincer, P. R. Shaver, E. Borgida, & J. A. Bargh (Eds.), *APA handbook of personality and social psychology, Volume 1: Attitudes and social cognition.* (pp. 387–414). American Psychological Association. https://doi.org/10.1037/14341-013

National Retail Federation. (2022). *NRF Reports Retail Shrink Nearly a $100B Problem*. https://nrf.com/media-center/press-releases/nrf-reports-retail-shrink-nearly-100b-problem

Nødtvedt, K. B., Sjåstad, H., Skard, S. R., Thorbjørnsen, H., & Van Bavel, J. J. (2021). Racial bias in the sharing economy and the role of trust and self-congruence. *Journal of Experimental Psychology: Applied*. https://doi.org/10.1037/xap0000355

OECD. (2017). *OECD Guidelines on Measuring Trust*. OECD. https://doi.org/10.1787/9789264278219-en

Pascual-Ezama, D., Fosgaard, T. R., Cardenas, J. C., Kujal, P., Veszteg, R., Gil-Gómez de Liaño, B., Gunia, B., Weichselbaumer, D., Hilken, K., Antinyan, A., Delnoij, J., Proestakis, A., Tira, M. D., Pratomo, Y., Jaber-López, T., & Brañas-Garza, P. (2015). Context-dependent cheating: Experimental evidence from 16 countries. *Journal of Economic Behavior & Organization*, *116*, 379–386. https://doi.org/10.1016/j.jebo.2015.04.020

Peer, E., Rothschild, D., Gordon, A., Evernden, Z., & Damer, E. (2021). Data quality of platforms and panels for online behavioral research. *Behavior Research Methods*, *54*(4), 1643–1662. https://doi.org/10.3758/s13428-021-01694-3

Pennycook, G., & Rand, D. G. (2022). Accuracy prompts are a replicable and generalizable approach for reducing the spread of misinformation. *Nature Communications*, *13*(1), 2333. https://doi.org/10.1038/s41467-022-30073-5

Perugini, M., Gallucci, M., Presaghi, F., & Ercolani, A. P. (2003). The personal norm of reciprocity. *European Journal of Personality*, *17*(4), 251–283. https://doi.org/10.1002/per.474

Pew Research Center. (2023). *In Divided Washington, Americans Have Highly Negative Views of Both Parties' Leaders*. https://www.pewresearch.org/politics/2023/04/07/evaluations-of-the-economy-and-the-state-of-the-nation/#:~:text=While%20Americans'%20views%20of%20current,think%20the%20economy%20will%20improve.

R Core Team. (2022). *R: A language and environment for statistical computing.* (4.1.1) [Computer software]. R Foundation for Statistical Computing. https://www.r-project.org/

Robbins, J. M., & Krueger, J. I. (2005). Social Projection to Ingroups and Outgroups: A Review and Meta-Analysis. *Personality and Social Psychology Review*, *9*(1), 32–47. https://doi.org/10.1207/s15327957pspr0901_3

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. https://doi.org/10.1016/j.joep.2014.10.002

Rotman, J. D., Khamitov, M., & Connors, S. (2018). Lie, Cheat, and Steal: How Harmful Brands Motivate Consumers to Act Unethically. *Journal of Consumer Psychology*, *28*(2), 353–361. https://doi.org/10.1002/jcpy.1002

Schilke, O., Reimann, M., & Cook, K. S. (2021). Trust in Social Relations. *Annual Review of Sociology*, *47*(1), 239–259. https://doi.org/10.1146/annurev-soc-082120-082850

Schotter, A., & Trevino, I. (2014). Belief Elicitation in the Laboratory. *Annual Review of Economics*, *6*(1), 103–128. https://doi.org/10.1146/annurev-economics-080213-040927

Schwarzer, G., Carpenter, J. R., & Rücker, G. (2015). *Meta-Analysis with R*. Springer International Publishing. https://doi.org/10.1007/978-3-319-21416-0

Sharot, T., Korn, C. W., & Dolan, R. J. (2011). How unrealistic optimism is maintained in the face of reality. *Nature Neuroscience*, *14*(11), 1475–1479. https://doi.org/10.1038/nn.2949

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-Positive Psychology: Undisclosed Flexibility in Data Collection and Analysis Allows Presenting Anything as Significant. *Psychological Science*, *22*(11), 1359–1366. https://doi.org/10.1177/0956797611417632

Skowronek, S. E. (2022). DENIAL: A Conceptual Framework to Improve Honesty Nudges. *Current Opinion in Psychology*, 101456. https://doi.org/10.1016/j.copsyc.2022.101456

Soraperra, I., Weisel, O., & Ploner, M. (2019). Is the victim *Max (Planck)* or *Moritz* ? How victim type and social value orientation affect dishonest behavior. *Journal of Behavioral Decision Making*, *32*(2), 168–178. https://doi.org/10.1002/bdm.2104

Sterling, T. D. (1959). Publication Decisions and their Possible Effects on Inferences Drawn from Tests of Significance—Or Vice Versa. *Journal of the American Statistical Association*, *54*(285), 30–34. https://doi.org/10.1080/01621459.1959.10501497

Sterling, T. D., Rosenbaum, W. L., & Weinkam, J. J. (1995). Publication Decisions Revisited: The Effect of the Outcome of Statistical Tests on the Decision to Publish and Vice Versa. *The American Statistician*, *49*(1), 108–112. https://doi.org/10.1080/00031305.1995.10476125

Sulitzeanu-Kenan, R., Tepe, M., & Yair, O. (2022). Public-Sector Honesty and Corruption: Field Evidence from 40 Countries. *Journal of Public Administration Research and Theory*, *32*(2), 310–325. https://doi.org/10.1093/jopart/muab033

Tarrant, M., Branscombe, N. R., Warner, R. H., & Weston, D. (2012). Social identity and perceptions of torture: It's moral when we do it. *Journal of Experimental Social Psychology*, *48*(2), 513–518. https://doi.org/10.1016/j.jesp.2011.10.017

Teper, R., Inzlicht, M., & Page-Gould, E. (2011). Are We More Moral Than We Think?: Exploring the Role of Affect in Moral Behavior and Moral Forecasting. *Psychological Science*, *22*(4), 553–558. https://doi.org/10.1177/0956797611402513

The jamovi project. (2022). *Jamovi.* (2.3) [Computer software]. https://www.jamovi.org

Vazire, S. (2018). Implications of the Credibility Revolution for Productivity, Creativity, and Progress. *Perspectives on Psychological Science*, *13*(4), 411–417. https://doi.org/10.1177/1745691617751884

Verrina, E., Galeotti, F., & Rilke, R. M. (2024). *Beliefs and Group Dishonesty: The Role of Strategic Interaction and Responsibility*. https://doi.org/10.2139/ssrn.4859462

Vincent, L. C., Emich, K. J., & Goncalo, J. A. (2013). Stretching the Moral Gray Zone: Positive Affect, Moral Disengagement, and Dishonesty. *Psychological Science*, *24*(4), 595–599. https://doi.org/10.1177/0956797612458806

Vosgerau, J., Simonsohn, U., Nelson, L. D., & Simmons, J. P. (2019). 99% impossible: A valid, or falsifiable, internal meta-analysis. *Journal of Experimental Psychology: General*, *148*(9), 1628–1639. https://doi.org/10.1037/xge0000663

Wang, L., & Murnighan, J. K. (2017). How Much Does Honesty Cost? Small Bonuses Can Motivate Ethical Behavior. *Management Science*, *63*(9), 2903–2914. https://doi.org/10.1287/mnsc.2016.2480

Weber, T. O., Schulz, J. F., Beranek, B., Lambarraa-Lehnhardt, F., & Gächter, S. (2023). The behavioral mechanisms of voluntary cooperation across culturally diverse societies: Evidence from the US, the UK, Morocco, and Turkey. *Journal of Economic Behavior & Organization*, *215*, 134–152. https://doi.org/10.1016/j.jebo.2023.09.006

Wiltermuth, S. S. (2011). Cheating more when the spoils are split. *Organizational Behavior and Human Decision Processes*, *115*(2), 157–168. https://doi.org/10.1016/j.obhdp.2010.10.001

Zhao, X., & Epley, N. (2022). Surprisingly Happy to Have Helped: Underestimating Prosociality Creates a Misplaced Barrier to Asking for Help. *Psychological Science*, *33*(10), 1708–1731. https://doi.org/10.1177/09567976221097615