Journal of Experimental Psychology: General

Lying to Appear Honest

Shoham Choshen-Hillel, Alex Shaw, and Eugene M. Caruso

Online First Publication, January 30, 2020. http://dx.doi.org/10.1037/xge0000737

CITATION
Choshen-Hillel, S., Shaw, A., & Caruso, E. M. (2020, January 30). Lying to Appear Honest. *Journal of Experimental Psychology: General*. Advance online publication. http://dx.doi.org/10.1037/xge0000737

American Psychological Association

Journal of Experimental Psychology: General

© 2020 American Psychological Association
ISSN: 0096-3445

http://dx.doi.org/10.1037/xge0000737

Lying to Appear Honest

Shoham Choshen-Hillel
The Hebrew University of Jerusalem

Alex Shaw
University of Chicago

Eugene M. Caruso
University of California, Los Angeles

People try to avoid appearing dishonest. Although efforts to avoid appearing dishonest can often reduce lying, we argue that, at times, the desire to appear honest can actually lead people to lie. We hypothesize that people may lie to appear honest in cases where the truth is highly favorable to them, such that telling the truth might make them appear dishonest to others. A series of studies provided robust evidence for our hypothesis. Lawyers, university students, and MTurk and Prolific participants said that they would have underreported extremely favorable outcomes in real-world scenarios (Studies 1a–1d). They did so to avoid appearing dishonest. Furthermore, in a novel behavioral paradigm involving a chance game with monetary prizes, participants who received in private a very large number of wins reported fewer wins than they received; they lied and incurred a monetary cost to avoid looking like liars (Studies 2a–2c). Finally, we show that people's concern that others would think that they have overreported is valid (Studies 3a–3b). We discuss our findings in relation to the literatures on dishonesty and on reputation.

Keywords: decision-making, lying, "behavioral ethics", social signaling, honesty

Supplemental materials: http://dx.doi.org/10.1037/xge0000737.supp

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.

Most people do not want to appear dishonest. In many cases, the best way for them to avoid appearing dishonest is to actually be honest (Akerlof, 1983). Indeed, even in the face of incentives for dishonesty, people routinely behave honestly, and are much more honest when their reputation for honesty is on the line (Gneezy, Kajackaite, & Sobel, 2018; Mazar, Amir, & Ariely, 2008). In this article, however, we demonstrate that, in predictable circumstances, a desire to appear honest can actually make people more likely to *lie*. Specifically, we examine situations in which reporting the true outcome is highly favorable to oneself, such that telling the truth might make one appear to be lying.

Imagine, for example, that your teacher lost all students' grades on an assignment and asked the students to honestly report their grades to him. Suppose the test was hard and many students barely passed, but you got a perfect score. You may worry that if you truthfully report your score, the teacher might think that you inflated it, so you may report that you got less than a perfect score. We hypothesize that in situations like these, people may be so concerned with appearing dishonest that they may lie and report a less favorable outcome. We start by reviewing previous research on people's motivations to lie and then report evidence that people sometimes lie to appear honest.

Lying

People's default is telling the truth (Grice, 1991). Telling a lie is generally viewed as morally wrong and is psychologically taxing (Gneezy, 2005). Thus, most people need a reason to tell a lie. At the same time, many situations provide such reasons, and people report being frequently dishonest in their day-to-day lives (DePaulo, Kashy, Kirkendol, Wyer, & Epstein, 1996; Hofmann, Wisneski, Brandt, & Skitka, 2014).

Why do people lie? Surely there are many reasons. One prominent motivation is gaining material benefits (Becker, 1968). Many people overcome their basic aversion to lying and mislead others to secure some monetary gains. People underreport income to pay less in taxes (Mazur & Plumley, 2007) and underreport mileage to

Shoham Choshen-Hillel, School of Business Administration and the Federmann Center for the Study of Rationality, The Hebrew University of Jerusalem; Alex Shaw, Department of Psychology, University of Chicago; Eugene M. Caruso, Anderson School of Management, University of California, Los Angeles.

We thank Emma Levine and Jane Risen for their constructive feedback on earlier versions of this article. We are also thankful to a team of excellent research assistants at the Hebrew University, led by Mika Guzikevits and Maya Enisman. We thank the Recanati Fund of the School of Business Administration at the Hebrew University and the Social Enterprise Initiative at Booth School of Business, The University of Chicago, for funding. Finally, Shoham Choshen-Hillel thanks her father, Dr. Ehud Choshen, Adv., for being the inspiration for Study 1a, as well as for being the inspiration for her research career in general.

The design, sample size, and analysis plans for Studies 1a, 1c, 1d, 3a, and 3b and for S1 in the online supplemental material were preregistered and can be accessed online (see links in the Methods of these studies). All data will be made available upon publication.

Correspondence concerning this article should be addressed to Shoham Choshen-Hillel, School of Business Administration and the Federmann Center for the Study of Rationality, The Hebrew University of Jerusalem, Mount Scopus, Jerusalem 9190501, Israel. E-mail: shoham@huji.ac.il

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.

insurance companies to reduce what they pay in premiums (Shu, Mazar, Gino, Ariely, & Bazerman, 2012). Abundant findings from experimental studies provide converging evidence demonstrating that many people lie to gain more money (e.g., Bryan, Adams, & Monin, 2013; Caruso & Gino, 2011; Gneezy, 2005; Mazar et al., 2008; Schurr & Ritov, 2016; Schurr, Ritov, Kareev, & Avrahami, 2012; for review, see Gerlach, Teodorescu, & Hertwig, 2019). The tendency to engage in profitable lies has been elegantly captured in the "die-under-the-cup" paradigm (Shalvi, Dana, Handgraaf, & De Dreu, 2011). Here, participants are asked to privately roll a die, knowing that they would be compensated according to the result they reported—the higher the roll, the more they are paid. Although people typically do not lie to the full extent (arguably because they want to maintain a moral self-image), on average, they do overreport their outcomes. Thus, a person who rolled a 3 in the die-under-the-cup paradigm may report a 5, to increase her gains (see also Cohn, Fehr, & Maréchal, 2014; Fischbacher & Föllmi-Heusi, 2013; Weisel & Shalvi, 2015). Indeed, in situations where deception is profitable, lying seems to be the default and self-control is required to overcome it (Gino, Schweitzer, Mead, & Ariely, 2011; Shalvi, Eldar, & Bereby-Meyer, 2012).

Yet not all lies are motivated by greed. Recent research has documented that people sometimes lie to help or to be kind to others. These lies, termed *prosocial lies*, involve telling things one knows to be untrue with the intention of helping someone else (Levine & Schweitzer, 2014). For example, a doctor may tell a dying patient that there is still hope that medication will help—even if she knows this is untrue—out of concern for the patient's emotional wellbeing (Levine et al., 2018). In addition, in controlled lab settings, participants have been found to tell a lie and forgo some monetary profit to enhance the profit of another participant in a study (Erat & Gneezy, 2012; Levine & Schweitzer, 2015).

In this article, we suggest a novel motivation for lying that is driven by neither monetary gains nor benevolence, but rather by concern with appearing dishonest. It may seem odd that we would suggest that wanting to avoid appearing dishonest could lead to *more* lying. Traditionally, people's desire to appear honest has been suggested to *limit* lying (Akerlof, 1983). Indeed, people care about the image that they present to others. People want to appear as good people—generous, fair, and moral (Andreoni & Bernheim, 2009; Dana, Cain, & Dawes, 2006; Shaw, Choshen-Hillel, & Caruso, 2018). For example, people tend to choose options that favor them over other people but would lie and say that they used an impartial device to make this choice, to avoid appearing unfair (Batson, Kobrynowicz, Dinnerstein, Kampf, & Wilson, 1997). People lie less when the chance that someone will find out that they are lying is higher (Gneezy et al., 2018). Thus, when participants' outcomes in the lab are observable (rather than completely private), the average outcome they report (for compensation) is lower. Further, concerns with appearing dishonest can motivate people even when others cannot verify that they are lying. People are less likely to lie to the full extent (i.e., report the highest possible outcome) when their true outcome is unobservable than when it is not (Gneezy et al., 2018). Presumably, this is because when the participants' true outcome is unobservable, reporting the highest possible outcome might appear suspicious or dishonest. For example, a person who rolled a 3 in private in the die-under-the-cup paradigm may avoid reporting a 6 (i.e., the most profitable outcome) because if he does, he may appear dishonest. Instead, he may report a 5—which is less profitable but will not hurt his honest appearance. Indeed, it has recently been suggested that experimental paradigms where high outcomes seem less plausible are associated with lower degrees of dishonesty (Gerlach et al., 2019).

We explore situations where a person who wants to appear honest may be tempted to lie; specifically, when the truth is so favorable that it may cause one to appear to be lying. Imagine a participant who is compensated for rolling higher numbers on a die and who is asked to report the outcomes to the experimenter who cannot verify what she rolled. Then imagine she rolled three 6s in a row. What would she tell the experimenter she rolled? Of course, if she has a moral inclination to tell the truth and desires to maintain a moral self-image (Gino, Ayal, & Ariely, 2013; Shalvi, Gino, Barkan, & Ayal, 2015) she should tell the truth. She should also tell the truth if she wants to just maximize her financial outcomes. But, we suspect she may feel hesitant to do so because telling the truth, in this case, will very likely make it appear to others like she is lying. We predict this concern can override other motivations for telling the truth, such that some people will feel compelled to lie.

Tentative evidence for "lying to appear honest" comes from a small study in which 19 university students—and 12 nuns—were asked to report the outcome of a private die roll (Utikal & Fischbacher, 2013). Unlike the students who selfishly overreported the outcome of the die, the nuns underreported it. We argue that this motivation to lie to appear honest not only drives the behavior of those who wish to appear as saints, but of ordinary people too.

The Current Studies

Our main hypothesis in this article is that some people will lie to their own disadvantage when they think that telling the truth may make them appear dishonest. We expect people to worry about appearing dishonest when their outcome is so favorable that it seems "too good to be true." Existing data sets from the lab cannot easily be used to test this argument, because cases of extremely favorable outcomes (e.g., rolling three 6s in a row) are, by definition, rare. Therefore, to test this argument, we designed a series of studies that confronted people with extremely favorable outcomes and examined their reactions. In all our studies, we manipulated participants' outcomes such that the outcomes were either standard (e.g., an average number of wins in a game of chance) or extremely favorable (e.g., a string of wins in a game of chance). The dependent measure was the distance between the outcomes that participants reported and their true outcomes. We predicted that participants will lie and underreport their outcomes more when their outcomes were extremely favorable than when they were standard.

Studies 1a–1d used realistic scenarios to test our hypothesis and the proposed mechanism. Study 1a asked lawyers to imagine that they have worked either a standard or an extreme number of hours, and then asked them what number of hours they would bill their client. We expected lawyers who were told they worked an extreme number of hours to underreport this number, in order to appear honest in the eyes of their clients. Studies 1b–1d further tested the proposed mechanism underlying people's decision to lie; specifically, whether they lie to their disadvantage because they

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.

are concerned with their appearance. In Studies 2a–2c, we designed a novel paradigm, involving many wins in a game of chance, to investigate actual behavior. We tested whether people would lie to appear honest even when doing so is costly to them. Finally, in Studies 3a and 3b, we tested whether people's concern with appearing dishonest when reporting extremely favorable outcomes accords with actual evaluations by others; that is, whether observers judge those who report an extremely favorable outcome more harshly than those who report a standard outcome, and whether underreporting extreme outcomes improves judgment.

Study 1

In Studies 1a–1d, we tested whether people would lie to appear honest, using realistic scenarios. In Study 1a, we examined whether (actual) lawyers would underreport to their clients an extremely high number of hours they worked. In Studies 1b–1c, we recruited both student and M-Turk populations, with two new scenarios, to test whether they would underreport extremely favorable outcomes and, additionally whether their concern with appearing dishonest would account for this tendency. Finally, in Study 1d we recruited British employees on Prolific and manipulated concern with appearing dishonest. We tested the effect on participants' likelihood to underreport extremely favorable outcomes.

Study 1a

In Study 1a, we tested whether people would lie to appear honest. As a first attempt to investigate this phenomenon, we recruited a sample of lawyers. We asked the lawyers to imagine that they worked on a case and were paid by the client according to the number of hours they reported. They either worked a standard number of hours, or an extremely high number of hours, and had to indicate the number of hours they would have reported to the client. We examined whether lawyers would underreport the number of hours worked when it was extremely high, but not when it was standard. If they did, then this would provide some initial evidence for the occurrence of the lying to appear honest phenomenon. We acknowledge that there could be alternative explanations for why lawyers may underreport a high number of hours even in this experimental design, which is why in Studies 1b–1d, we test our hypothesis with more tightly controlled scenario studies and directly test whether concern with appearance accounts for this effect.

Method.

Participants. Participants were lawyers who have passed the Israeli bar exams. The lawyers were recruited through Facebook groups restricted to lawyers, collective emails sent through contact people at law firms, and a personal Facebook ad that invited lawyers to send us an e-mail and receive a Qualtrics link to the study. One hundred and fifteen lawyers participated in this 5-min study, contributing to science and entering a lottery for restaurant gift cards. One hundred and nine participants completed the demographic questions (50.5% females, age ranged from 26 to 66, $M_{age} = 35.28$, $SD_{age} = 5.70$). The lawyers varied in their legal experience (they passed their bar exams between the year they completed the study and 23 years before $M_{bar} = 7.83$, $SD_{bar} = 4.48$). We preregistered this study on https://aspredicted.org/he4t3.pdf. According to our preregistered rule, we stopped data collection a week after recruitment started. We used all the responses that were collected. Ethics approval for this and for all the following studies was obtained from the university's institutional review board.

Procedure. Lawyers were invited to participate in a "decision making study." In neither this nor any of the following studies were participants told the true purpose of the study, and there was no mention that it was about "honesty" or "morality." The participants read the following scenario (in Hebrew):

> Imagine that you are working on a case for a client who pays you privately on an hourly basis.
>
> You work from the office so the client cannot tell what number of hours you truly put into the case.
>
> Before you took the case, you had estimated that you would need between 60 and 90 working hours, but you explained to the client that the ultimate number would depend on the amount of work that would be needed.

Participants were randomly assigned into one of two conditions—extreme or standard outcome condition. Participants in the extreme outcome condition read: "You ended up working on the case 90 hours." Participants in the standard outcome condition instead read: "You ended up working on the case 60 hours." All participants had to type in the number of hours they would bill the client (in a free text box). Participants were then asked to explain their answer in an open-ended question. Finally, they were asked a few demographic questions, and how many years ago they had passed the bar exams. Participants were invited (but not required) to leave their e-mail address to participate in the restaurant gift cards lottery. We assured them that e-mail addresses would be kept separately from the data, to ensure the anonymity of their responses.

Results.

Lying. For each participant, we computed a "lying score" by subtracting the actual number of hours she worked (60 in the standard outcome condition and 90 in the extreme outcome condition) from the number of hours she indicated she would bill the client. A t test revealed that the lying score was indeed lower in the extreme outcome condition ($M = -1.92$, $SD = 4.72$) than in the standard outcome condition ($M = 2.55$, $SD = 6.86$), $t(113) = 4.05$, $p < .001$, $d = 0.76$. Specifically, whereas participants in the extreme outcome condition reported a number of hours that was significantly lower than the truth (i.e., their lying scores were lower than 0), $t(55) = 3.04$, $p = .004$, $d = 0.41$, participants in the standard outcome condition reported a number of hours that was significantly higher than the truth (i.e., their lying scores were higher than 0), $t(59) = 2.86$, $p = .006$, $d = 0.37$. In the standard outcome condition ($n = 59$), 17% of the lying scores were greater than zero, 83% were equal to zero, and none were smaller than zero (0%), indicating that participants either told the truth or overreported. In contrast, in the extreme outcome condition ($n = 56$), 0% of the lying scores were greater than zero, 82% were equal to zero, and 18% were smaller than zero, indicating that participants either told the truth or underreported.

Participants' explanations. Anecdotally, some participants in the extreme condition explained their decision to underreport the

high number of hours they had worked by referring to concern with appearing dishonest (e.g., ". . . Billing 90 hours would appear bad because the client would think I cheated him with the hours . . ."). We coded the explanations of the participants who underreported their outcomes in the extreme outcome condition (n = 10), to see the prevalence of different motivations they brought up. Two independent judges read the texts of the explanations only and were instructed to indicate the presence (1) or absence (0) of each of the following motivations: (a) concern with appearing dishonest, (b) concern with getting caught, (c) concern with appearing incompetent, and (d) concern with fairness. An explanation could involve, in principle, between zero and four motivations. The interjudge correlation was .85. For concern with appearing dishonest, one judge indicated the explanation reflected this motivation in 50% of responses, and the other indicated it in 40% of responses. Fifty percent of the explanations were rated by at least one judge to reflect a concern with appearing dishonest. For concern with appearing incompetent, one judge indicated the explanation reflected this motivation in 40% of responses, and the other indicated it in 30% of responses. Forty percent of the explanations were rated by at least one judge to reflect a concern with appearing dishonest. Both judges indicated that none of the explanations reflected concern with getting caught or concern with fairness. Thus, the open-ended explanations provided anecdotal evidence that at least some of the lawyers who underreported did so because they were concerned about appearing dishonest. Yet others may have underreported because they were worried that if they reported it, the client would think that they are slow or incompetent.

Discussion. According to the findings of Study 1a, some lawyers indicated that they would have underreported the number of hours they worked on a case had this number been extremely high (but not standard). According to our account, the reason lawyers would underreport extreme outcomes is that they are worried that the client would suspect that they overreported the number of hours they truly worked in order to earn more. Whereas Study 1a suggested that this concern may have driven some of the lawyers who had underreported the high number of hours, other motivations could have also accounted for their behavior, and, in particular, concern with appearing incompetent. Studies 1b–1d were designed to examine if we could observe underreporting in contexts where it would not have the added benefit of making one appear competent, and to test our suggested mechanism more directly.

Study 1b

In Study 1a, lawyers indicated that they would bill their client for fewer hours than they actually worked on a case, if the number of hours they worked was very high. Study 1b was designed to rule out an alternative explanation, whereby this tendency to underreport was driven by concern with appearing competent, rather than by concern with appearing honest.

Specifically, in Study 1b, similar to our opening example, university students were asked to imagine that their TA did not have access to their grades, and asked them to self-report their grades—which were either average or extremely good (this actually happened to one of the authors of this article). We predicted that students would underreport extremely good grades more so than average grades. If students underreport good grades, then they are clearly not trying to appear competent. Instead, according to our suggested theory, they underreport because they are trying to appear honest. Thus, we expected students' reported concern with appearing dishonest to mediate underreporting.

Method.

Participants. One hundred fifty-nine undergraduate students in an Israeli university participated in this study (59% females, M_{age} = 28.55, SD_{age} = 5.68). Participants were recruited through a lab mailing list and were invited to complete a short online survey in exchange for participation in a lottery (they were promised a chance to win one of three prizes of 100 shekels, each equivalent to about $30 at the time of the study). According to a predetermined rule, we excluded from further analyses all participants who did not answer correctly the reading comprehension question. Five participants were excluded based on this rule.

Procedure. The participants read the following scenario (in Hebrew):

> Imagine that you are a student at a class where there are 12 mandatory exercises. The TA grades each exercise as either "fail," "pass," or "excellent." To take the final exam, a student must obtain a pass on at least 10 exercises. Other than that, the grades of the exercises do not count toward the final grade in the course. At the end of the semester, the TA suddenly has to travel abroad. Because he does not have access to the exercises, he asks you to go to his drawer, collect the exercises, and type the grades into the Excel sheet, and send it to him.

Participants were randomly assigned into one of two conditions—extreme outcome or standard outcome condition. Participants in the extreme outcome condition read, "While you are typing, you notice that your grades are exceptional—not only did you pass all exercises, you got 'Excellent' on all of them. Most of the students received between 4 to 6 Excellent grades." In the standard outcome condition, participants instead read, "While you are typing, you notice that your grades are standard—you passed all exercises, and got 'Excellent' on five of them. Most of the students received between 4 to 6 Excellent grades." Participants in both conditions were then asked, "How many excellent grades would you type in?" They had to indicate their response on a numeric scale that ranged from 1 to 12 and then explain it in an open-ended question (specifically, to answer the question "Why?"). The participants were then asked, "If you type in your actual number of excellent grades, how likely is the TA to think that you are a dishonest person?" Participants rated their answer on a 1 (*not likely at all*) to 7 (*very likely*) scale. Participants were then asked one comprehension question ("What was the true number of Excellent grades you received?") and a few demographic questions.

Results.

Lying. For each participant, we computed a "lying score" by subtracting the actual number of "excellent" grades the participant got from the number of "excellent" grades she reported. A t test revealed that participants' lying score was lower when they found out they had perfect grades (extreme outcome condition, $M = -0.27$, $SD = 1.11$) than when they found out they had standard grades (standard outcome condition, $M = 0.01$, $SD = 0.11$), $t(154) = 2.25$, $p = .026$, $d = 0.36$. Specifically, whereas in the extreme grades condition participants significantly underreported their outcomes (i.e., their lying score was lower than zero),

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.

$t(77) = 2.16$, $p = .034$, $d = 0.24$, in the standard outcome condition participants' reports were not significantly different from the truth, $t(77) = 1.00$, $p = .320$. Participants in the extreme outcome condition were more likely to underreport their outcome (8%) than participants in the standard outcome condition (0%), $p = .028$ ($N = 154$), Fisher's exact test, two-tailed.

Process measures. According to our reasoning, participants in the extreme outcome condition underreported their outcome because they feared that the truth might make them appear dishonest. Indeed, participants in the extreme outcome condition indicated that if they reported the truth, the TA was more likely to think that they were being dishonest ($M = 4.67$, $SD = 1.90$) than participants in the standard outcome condition ($M = 3.00$, $SD = 2.06$), $t(154) = 5.26$, $p < .001$, $d = 0.85$. Concern with appearing dishonest was correlated with the lying score in the extreme outcome condition (i.e., the more participants were concerned with appearing dishonest, the more they underreported), $r = -.24$, $p = .037$. In the standard outcome condition, this correlation was not significant, $r = .06$, $p = .628$. To test whether the participants' outcomes affected their tendency to underreport their outcomes through their concern with appearing dishonest, we conducted a mediation analysis. The experimental condition of outcome had a significant effect on the lying score ($b = -0.28$, $p = .026$), and condition also affected concern with appearing dishonest ($b = 1.67$, $p < .001$). The lying score was directionally, but not significantly, affected by the concern with appearing dishonest ($b = -0.06$, $p = .054$). Condition was not a significant predictor of the lying score when the mediator was entered into the analysis ($b = -0.18$, $p = .185$), suggesting mediation. The 95% bias-corrected confidence interval for the size of the total indirect effect of condition on lying through concern with appearing dishonest excluded zero [−0.34, −0.01], suggesting significant mediation.

Participants' explanations. Anecdotally, some participants in the extreme outcome condition explained their decision to underreport by expressing their concern with appearing dishonest. For example, one participant wrote, "If the TA does not remember the grades, he might think I took advantage of the opportunity to type in the grades," and another wrote, "Because it appears too good to be true." We coded the explanations given by participants in the extreme condition who underreported their outcomes ($n = 6$) in the same manner as in Study 1a. The interjudge correlation was .77. For concern with appearing dishonest, one judge indicated the explanation reflected this motivation in 100% of responses, and the other indicated it in 67% of responses. One hundred percent of the explanations were rated by at least one judge to reflect a concern with appearing dishonest. Both judges indicated that none of the explanations reflected concern with appearing incompetent, concern with getting caught, or concern with fairness (i.e., they were never mentioned). Thus, participants' explanations provided some additional, anecdotal evidence that it was concern with appearing dishonest, rather than other motivations, that led them to underreport their extreme outcomes.

Discussion. Study 1b provided support for our hypothesis that participants would lie to appear honest and that their concern with appearing dishonest accounted for their tendency to lie in such cases. Here, university students said they would have underreported an extremely high set of scores, even though these scores were based on their true skills. These results rule out the possibility that participants are underreporting their favorable outcomes to appear more competent.

Study 1c

In Study 1c, we sought to generalize the findings from Studies 1a–1b. We now recruited participants on MTurk and used a new organizational scenario where an employee had to report to their boss their monthly mileage in work trips for monetary compensation. The mileage report was unrelated to the employee's skills or competence. Employees drove either a standard or an extremely high number of miles. We predicted again that participants would underreport extreme (but not standard) mileage. According to our theory, this tendency to underreport is driven by participants' concern with appearing dishonest when reporting extremely favorable outcomes. To test this, we measured participants' concern with appearing dishonest (if they told the truth). We predicted that this concern would mediate the tendency to lie to appear honest.

Method.

Participants. Participants were recruited online using the Amazon Mechanical Turk platform; participation was restricted to adult participants from the United States. Two hundred and one adults (49% females, $M_{age} = 36.38$, $SD_{age} = 12.96$) participated in this 5-min study for 25 cents. We preregistered our hypothesis and sample size on https://aspredicted.org/h9ue5.pdf. A sample size calculation based on previous results (see Study 2a and 2b that chronologically preceded Study 1c), where we obtained medium effect sizes (equivalent to $d = 0.5$), revealed that to achieve 0.8 power, we should aim for 64 participants per cell. We aimed for 100 participants per cell, allowing for exclusion of participants who fail the comprehension question (around 10%), and greater power (.9). According to the preregistered rule, we excluded 17 participants who failed the reading comprehension question.

Procedure. The participants read the following scenario:

> Imagine that you are working for a large company. Your job includes many work trips in your car, and the company compensates you for the mileage.
>
> The company's policy is that at the end of each month, each employee needs to tell their manager the number of miles they drove that month, and they then get paid accordingly. The more miles you report, the more compensation you receive. The **maximum** that the company would compensate for is **400** miles.
>
> Many employees go on similar work trips, **normally driving between 280 and 320 miles per month**.

Participants were randomly assigned into one of two conditions—extreme or standard outcome condition. Participants in the extreme outcome condition read, "This month, many of your work trips were plagued by detours because of unannounced construction projects, and so **you traveled 400 miles**." In the standard outcome group, participants instead read, "This month, your work trips went as usual, and so **you traveled 300 miles**." All participants then had to indicate their response to the question: "How many miles would you report to your manager?" on a slider scale that ranged from 0 to 400 miles. The participants were also asked to rate on a 1 (*not likely at all*) to 7 (*very likely*) scale: "If you report the actual number of miles that you drove, how likely is the manager to think that you are being dishonest?" They also had to

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.

explain their answer in an open-ended question. Participants were asked one comprehension question ("How many miles did you drive this month?") and a few demographic questions.

Results.

Lying. For each participant, we computed a "lying score" by subtracting the actual number of miles she drove from the number of miles she reported. A *t* test revealed that participants' lying score was lower when they found out they traveled the maximum number of miles (extreme outcome condition, $M = -16.15$, $SD = 57.81$) than when they found out they traveled a standard number of miles (standard outcome condition, $M = 1.14$, $SD = 22.71$), $t(182) = 2.62$, $p = .009$, $d = 0.39$. Specifically, whereas participants in the extreme outcome condition reported a number of miles that was significantly lower than the truth, $t(92) = 2.74$, $p = .007$, $d = -0.28$, participants in the standard outcome condition reported a number of miles that was not significantly different from the truth, $t(87) = 0.47$, $p = .640$. In addition, participants in the extreme outcome condition were directionally more likely to underreport their outcome (11.5%) than participants in the standard outcome condition (4.5%), but this difference was not significant, $\chi^2(N = 184) = 4.23$, $p = .087$, $\varphi = 0.13$.

Process measures. According to our argument, participants underreported their outcomes when they drove many miles because they feared that reporting the truth might make them appear dishonest. Indeed, participants in the extreme outcome condition indicated that the manager was more likely to think that they were being dishonest ($M = 3.63$, $SD = 2.07$) than participants in the standard outcome condition ($M = 2.59$, $SD = 1.83$), $t(182) = 3.58$, $p < .001$, $d = 0.53$. Anecdotally, a participant in the extreme outcome condition explained the decision to report 350 instead of 400 miles by saying "If most employees are reporting mileage within a certain range, than [sic] a mileage report that is much higher will look suspicious" (see Study 1d for an analysis of participants' explanations in Studies 1c and 1d). Concern with appearing dishonest was significantly correlated with the lying score in the extreme outcome condition (i.e., the more participants were concerned with appearing dishonest, the more they underreported), $r = -.21$, $p = .036$. The correlation was not significant in the standard outcome condition, $r = .04$, $p = .718$.

To test whether the participants' outcomes affected their tendency to underreport through their concern with appearing dishonest, we conducted a bootstrapping mediation analysis (Hayes, 2013). The experimental condition of outcome had a significant effect on the lying score ($b = -18.42$, $p = .006$). Outcome also affected participants' concern with appearing dishonest ($b = 1.03$, $p < .001$). The lying score was directionally, but not significantly, affected by the concern with appearing dishonest ($b = -2.79$, $p = .098$). Condition was still a significant predictor of lying when the mediator was entered into the analysis ($b = -15.53$, $p = .023$), suggesting partial mediation. The 95% bias-corrected confidence interval for the size of the total indirect effect of condition on distance from the truth through concern with appearing dishonest excluded zero [−8.16, −0.18], suggesting a significant mediation.

Discussion. In Study 1c, we found once again that participants whose outcomes seemed "too good to be true" (i.e., they drove the maximum mileage for which their company was willing to compensate) underreported their outcomes more than participants whose outcomes were standard (i.e., they drove the average number of miles). This greater tendency to underreport in the extreme outcome condition was partially explained by participants' greater concern with appearing dishonest in this condition, in case of telling the truth. In our supplemental materials, we report a direct replication of this result (see Supplemental Study S1 in the online supplemental material), using a scale that allowed for overreporting in both conditions, and investigating individual differences in the tendency to underreport. The main findings of both these studies support our hypothesis that people may lie to avoid appearing as liars.

Study 1d

In Studies 1a–1c, we established that participants underreported extremely favorable outcomes. According to our theory, they do so because they are worried that if they report the truth, they would appear dishonest; that is, others would think that they had overreported their outcomes to gain more money. Indeed, we found that concern with appearing dishonest mediated the tendency to underreport. In Study 1d, we aimed to establish a causal relationship, and tested whether concern with appearing dishonest also moderated the effect.

To test whether people are less likely to underreport their extreme outcomes when they are not concerned that they will appear dishonest, we used the extreme and standard outcome conditions from Study 1c's mileage scenario. As in Study 1c, participants either read that they drove an extreme number of miles, or a standard one, and were asked to report to their boss the number of miles for compensation. We added two new conditions. The new conditions were similar to the ones used in Study 1c, except that participants were told that an automatic system would send a message to their boss if their report was *higher* than the actual mileage they drove. Thus, the manager would know if they lied to their advantage or not. This meant that participants should not be worried about appearing as liars when reporting the truth (i.e., the manager would know they have not overreported). If participants are reluctant to report extreme outcomes for reasons other than what we have suggested (e.g., because they think it is unfair toward the company to take so much money), then they should still underreport in this new extreme outcome condition. Note that they can still underreport in this scenario, because only overreporting, and not underreporting, is flagged by the automatic system. However, if our theory is correct, and participants underreport because they are concerned with appearing dishonest, then they should be less likely to underreport their extreme outcomes in this new condition where appearing as a person who overreported is not a concern anymore.

Method.

Participants. Participants were recruited online using the Prolific platform; participation was restricted to adult participants from the United Kingdom. Five hundred forty-four adults (70% females, $M_{age} = 35.75$, $SD_{age} = 12.10$) participated in this 5-min study for 0.35 GBP. We preregistered our hypothesis and sample size on https://aspredicted.org/rf44g.pdf. The sample size was determined according to a power analysis, whereby in order to

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.

achieve 0.8 power, for an effect of the size of 0.15[1] we should aim for 123 participants per cell. We aimed for 135 participants per cell (540 in total), allowing for exclusion of participants who fail the comprehension question (around 10%). According to the preregistered rule, we excluded 34 participants who failed at least one of the two reading comprehension questions.

Procedure. The participants read the following scenario (similar to the one used in Study 1c):

> Imagine that you are working for a large company. Your job includes many work trips in your car, and the company compensates you for the mileage.
>
> The company's policy is that at the end of each month, each employee needs to tell their manager the number of kilometers they drove that month, and they then get paid accordingly. The more kilometers you report, the more compensation you receive. The **maximum** that the company would compensate for is **400** km, so you are asked to report a number between 0 and 400 km.

The participants were randomly assigned into one of four conditions, according to a 2 × 2 design with outcome (extreme or standard) and overreporting (revealed or not) as between-participants factors. Only participants in the overreporting revealed conditions read: "Within a few days, an automatic system checks the report and sends the manager a message if it is higher than the car's actual mileage. Thus, your manager will know if you reported more kilometers than you actually drove."

All participants then read "Many employees go on similar work trips, **normally driving between 280 and 320 km** per month." Participants in the standard conditions further read: "This month, your work trips went as usual, and so **you traveled 300 km**."

Participants in the extreme conditions read instead: "This month, many of your work trips were plagued by detours because of unannounced construction projects, and so **you traveled 400 km.** You are the only employee who had to take these roads."

All the participants were asked, "How many kilometers would you report to your manager?" They had to indicate their response (in digits only) in an open-ended box. In the over-reporting-not revealed conditions, they were reminded before they answered: "Your manager will not know whether or not your report is higher than your actual mileage." In the overreporting revealed conditions, they were reminded instead: "Your manager will know if your report is higher than your actual mileage." On the next screen, the participants were asked, "Why did you choose to report that number of kilometers?" Participants were asked two comprehension questions ("How many kilometers did you drive this month?" "Would your manager be able to verify if you reported more kilometers than you actually drove?") and a few demographic questions.

Results.

Lying. First, we computed for each participant her lying score (the number of kilometers she reported minus the actual number she drove). To compare lying in the different conditions, we conducted a two-way analysis of variance (ANOVA) with outcome (extreme or standard) and overreporting (revealed or not) as between-participants factors. The dependent measure was the lying score. Consistent with our hypothesis, participants' lying score was lower in the extreme outcome conditions ($M = -1.31$, $SD = 7.86$) than in the standard outcome conditions ($M = 2.12$, $SD = 11.10$), $F(1, 506) = 18.45$, $p < .001$, $\eta_p^2 = 0.04$. There was no significant main effect for revealed overreporting, $F(1, 506) = 2.37$, $p = .124$. A significant interaction occurred between outcome and revealed overreporting, $F(1, 506) = 15.89$, $p < .001$, $\eta_p^2 = 0.03$. Specifically, when overreporting was not revealed, participants' lying score was lower in the extreme than in the standard outcome condition, $F(1, 506) = 33.05$, $p < .001$, $\eta_p^2 = 0.06$. By contrast, when overreporting was revealed, there was no significant difference in participants' lying scores between the extreme and standard outcomes conditions, $F(1, 506) = 0.05$, $p = .824$. Thus, in line with our prediction, we obtained the lying effect when overreporting could not be revealed (and one might appear as a liar), but not when overreporting was revealed (and there was no concern of appearing as a liar when reporting the truth). We also compared the lying scores between the two extreme outcome conditions, when overreporting was revealed or not. This contrast was not statistically significant, $F(1, 506) = 2.95$, $p = .087$, $\eta_p^2 = 0.01$.

Next, we examined in each condition separately whether the average lying score was higher, lower, or equal to zero (i.e., the truth). When overreporting was not revealed and the outcome was extreme, participants reported a number of kilometers that was significantly lower than the truth ($M = -2.32$, $SD = 10.78$), $t(126) = -2.43$, $p = .017$, $d = 0.23$. Yet, when overreporting was not revealed and the outcome was standard, the number of kilometers reported was significantly *higher* than the truth ($M = 4.66$, $SD = 16.16$), $t(117) = 3.13$, $p = .002$, $d = 0.29$. When overreporting was revealed, participants' reports were not significantly different from the truth, both in the extreme outcome condition ($M = 0.26$, $SD = 2.01$), $t(122) = -1.43$, $p = .154$, and in the standard outcome condition (where the mean and standard deviation were 0, indicating that none of the participants lied, and not allowing a statistical test). This analysis supports our contention that participants underreport extreme outcomes—but only when overreporting cannot be detected.

Next, we examined the rate of participants who underreported their outcome. A logistic regression could not be performed, because there was zero variance in the standard outcome condition where overreporting was revealed. We therefore compared the rate of participants who underreported their outcome in the extreme and standard outcomes conditions. When overreporting was not revealed, participants in the extreme outcome condition were more likely to underreport their outcome (6%) than participants in the standard outcome condition (0%), $p = .015$ ($N = 245$), Fisher's exact test, two-tailed. Also, when overreporting was revealed, participants in the extreme outcome condition were more likely to underreport their outcome (3%) than participants in the standard outcome condition (0%), $p = .045$ ($N = 265$).

Participants' explanations. Finally, we analyzed the explanations given by participants in Study 1c and 1d who had underreported their outcomes in the extreme outcome, overreporting not-

[1] The effect size in Study 1b was $d = 0.42$, but here we used moderation and wanted to allow detection of a smaller effect.

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.

revealed condition ($n = 18$). We coded the explanations in the same manner as in Studies 1a–1b. The interjudge correlation was .79. For concern with appearing dishonest, one judge indicated the explanation reflected this motivation in 67% of responses, and the other indicated it in 45% of responses. Sixty-seven percent of the explanations were rated by at least one judge to reflect a concern with appearing dishonest. Both judges indicated that none of the explanations reflected concern with appearing incompetent, concern with getting caught, or concern with fairness or standing out among other employees. Thus, once again, participants' spontaneous explanations for underreporting their extreme outcomes seemed to arise only from concern with appearing dishonest, rather than other motivations.

Discussion. We found that participants were much less likely to lie about their mileage in both the standard and extreme cases when their boss could verify if they overreported it. This result is obvious for the standard condition because any attempt to overreport would be immediately discovered as a lie. However, for the extreme condition, participants could have underreported without being detected (they were told that the boss would only be notified about overreporting). Still, participants lied and underreported less in this condition as compared to the extreme condition where their boss could not verify their report. We argue that we obtained this result because in this condition the employee does not need to worry that if they report the maximum (extreme) mileage they actually drove, their manager will view them as liars because they and their manager have common knowledge that over reporting is flagged by the system automatically. Note this reduction in underreporting is not due to the fact that participants would be more worried about being caught underreporting—the system only flags overreporting. Thus, these findings provide evidence that participants' underreporting was being driven by a concern with appearing as liars rather than by alternative motivations.

Studies 1a–1d provide evidence that people will lie to appear honest. However, these studies were based on participants' self-report, and it is possible that participants say they would have underreported their outcomes whereas in fact they would not have, due to the cost associated with underreporting. In Studies 2a–2c, we examine people's actual behavior in situations where lying to appear honest entails a monetary cost.

Study 2

In Studies 2a–2c, we designed a novel behavioral paradigm to test whether people would lie to appear honest when underreporting was costly to them. In Study 2a, we compared underreporting in two basic conditions (extreme and random outcomes). In Studies 2b–2c, we tested whether increased concern with appearing dishonest affected underreporting.

Study 2a

In Study 2a, we tested whether participants in a lab experiment would lie to avoid appearing dishonest, even though lying was costly to them. Participants had to roll a die four times and flip a coin eight times on Google's applications, and report how many "wins" they got. They were also told they would receive bonus (0.5 NIS) for each of these wins. Unbeknownst to them, we manipulated the program to control their outcomes, such that one group got 12 out of 12 wins, whereas another got a random result. We also measured participants' concern with appearing dishonest in the eyes of the experimenter (to whom they reported their wins). As in the previous studies, we predicted that participants would underreport more in the extreme than in the random condition, and that this difference would be driven by their concern with appearing dishonest.

Method.

Participants. One hundred forty-nine undergraduate students in an Israeli university (60% females, $M_{age} = 24.11$, $SD_{age} = 3.67$) participated in this study. The participants received a basic payment of 10 NIS (equivalent to about $3), and also received a bonus as explained below. A power analysis revealed that to achieve 0.8 power and detect a medium size effect, we should aim for 64 participants per cell. We aimed for 75 participants per cell to allow for technical errors. The data of 11 participants were not recorded due to technical errors in the program. One more participant was excluded from further analysis because they failed to follow the RA's instructions.

Procedure. The experimenter led each participant into a private room with a computer. The experimenter read the instructions to each participant personally, left the room during each task, and returned once the participant had completed the task. The participant was told that the goal of the study was to map users' experience with different Google applications. The participants were told that the study consisted of four tasks, in each of which she would be asked to use a different Google feature and report her experience. We used Google so that participants would believe the outcomes presented to them were authentic and not controlled by the experimenter. In the first task, the participant was instructed to open a Google browser and type in "Pacman," which led her to a standard Pacman game hosted by Google. The participant was asked to count to 10 once the experimenter left the room and then start playing, while keeping track of the number of ghosts she ate during the game. She was instructed to close the browser once the experimenter knocks on the door. The experimenter left the room and then knocked on the door and reentered after 3 min. The experimenter asked the participant to report to her the number of ghosts she ate, which the experimenter wrote down on a form. This procedure signaled to the participant that her results on Google remained private, and that the experimenter relied on her report. The experimenter also asked the participant to report how long she thought she had played Pacman. In the second part of the study, to further convince the participant that she was using Google, she was asked to engage in a free search on Google, while the experimenter left the room. When the experimenter returned, the participant was asked how long she thought she had spent in the search.

The next two tasks manipulated participants' outcomes and measured their tendency to lie. They followed the same procedure as the first two tasks (where the experimenter left the room during the task and then reentered). Participants were assigned to one of two conditions, extreme or random outcomes. In the third task, participants were asked to roll a die four times on Google's "roll-a-die" application and report the number of 5s and 6s they got. Participants were told that they would earn a 0.5 NIS bonus

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.

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.

(about 15 U.S. cents) for each 5 or 6 they flipped (in addition to their flat fee for participation).

Unbeknownst to participants, we wrote a computer program that allowed us to manipulate and record their die outcomes.[2] In the extreme outcome condition, the program always presented participants with a string of winning rolls (i.e., 5, 6, 6, 5).[3] In the random outcome condition, the program selected the outcomes of the die at random (e.g., 3, 3, 5, 1). The outcomes in the extreme outcome condition were designed to seem statistically probable, but to make participants concerned about appearing dishonest in the next task, when they get even more wins. When the experimenter reentered the room, she asked the participants to report how many times they got 5 or 6 in her four die rolls and wrote down this number. She then asked them to engage in another free search on Google and to estimate the time it took them.

In the last task, participants were asked to flip a coin eight times on Google's "flip-a-coin" application, and once the experimenter returned, report to them the number of heads they got. Participants were promised 0.5 NIS for each heads they reported. The program presented participants with a string of eight heads (extreme outcome condition)[4] or with a random sequence of heads and tails (random outcome condition).

Once participants reported their number of heads they received, they were given a questionnaire in which they were asked to indicate, "How concerned were you that if you reported your true dice outcomes, the experimenter would not believe you?" They were then asked the same question about their coin flip outcomes. For both questions, they had to indicate their response on a 1 (*not concerned at all*) to 7 (*highly concerned*) scale. Finally, the participants filled out some demographical questions, and were debriefed.

Because in the extreme outcome condition the favorable outcome of 12 wins out of 12 was statistically unlikely, we expected participants to be concerned that if they reported their actual number of wins, they would appear as liars. Because of this concern, we expected to see some underreporting of extreme outcomes, and less underreporting of random outcomes.

Results.

Lying. For each participant, we computed a "lying score" by subtracting the actual number of wins (total of 5s and 6s in the die rolls and heads in the coin flips, as recorded by the program) from the number of wins she reported. A t test revealed that participants' lying score in the extreme outcome condition ($M = -0.32$, $SD = 0.63$) was lower than in the random outcome condition ($M = 0.12$, $SD = 0.75$), $t(135) = 3.76$, $p < .001$, $d = 0.65$. Specifically, in the extreme outcome condition, when participants had 12 out of 12 wins, they significantly underreported their number of wins (i.e., their average lying score was lower than zero), $t(70) = 4.35$, $p < .001$, $d = 0.51$. In contrast, in the random outcome condition, when participants had a random number of wins, their average lying score was not significantly different from zero, $t(65) = 1.31$, $p = .197$. Consistent with these findings, participants in the extreme outcome condition were more likely to underreport their number of wins (24%) than participants in the random outcome condition (4.5%), $\chi^2(N = 137) = 10.32$, $p = .001$, $\varphi = .28$.

Process measures. According to our reasoning, participants in the extreme outcome condition underreported their number of wins because they feared that reporting the true number might make them appear dishonest in the eyes of the experimenter. For each participant, we computed a "concern with appearing dishonest" score, by averaging the participant's concern with appearing dishonest when reporting the die roll's true outcome, and when reporting the coin flip's true outcome. Indeed, participants in the extreme outcome condition had a higher concern with appearing dishonest score ($M = 4.50$, $SD = 1.64$) than participants in the random outcome condition ($M = 1.81$, $SD = 1.10$), $t(135) = 11.14$, $p < .001$, $d = 1.92$. Concern with appearing dishonest was significantly correlated with the lying score in the extreme outcome condition (i.e., the more participants were concerned with appearing dishonest, the more they underreported), $r = -.27$, $p = .022$. In the random outcome condition, this correlation was not significant, $r = -.02$, $p = .878$.

To test whether the participants' outcomes affected their tendency to underreport through their concern with appearing dishonest, we conducted a mediation analysis. The experimental condition of outcome had a significant effect on the lying score ($b = 0.45$, $p < .001$) and outcome also affected concern with appearing dishonest ($b = -2.68$, $p < .001$). The lying score was directionally, but not significantly, affected by the concern with appearing dishonest ($b = -0.07$, $p = .069$). Condition was not a significant predictor of the lying score when the mediator was entered into the analysis ($b = 0.23$, $p = .143$), suggesting mediation. The 95% bias-corrected confidence interval for the size of the total indirect effect of condition on lying through concern with appearing dishonest excluded zero [0.07, 0.38], suggesting significant mediation.

Discussion. Study 2a provided a replication of our basic effect in a consequential game of chance. We found that participants tended to lie and underreport an extremely large number of wins (which could seem "too good to be true"), even though underreporting was costly to them. Mediation analysis supported our argument that participants acted against their monetary incentives because they worried that the truth would make them appear to be liars.

Study 2b

Study 2a provided behavioral evidence for our hypothesis, whereby people would lie to appear honest. We found that participants who had an extremely high number of wins reported a lower number of wins, even though this hurt their pay. An alternative explanation for this finding, however, could be that participants underreported their extremely favorable outcomes because they miscounted their number of wins. To rule out this explanation, in Study 2b we replicated the extreme and standard outcome conditions and added two new conditions. In the new conditions, the

[2] The program was programed using JavaScript. It was installed as an add-on to Chrome on the lab computer before the experiment began and was not visible to the participants. The program was activated once the participants entered the Google roll-a-die or flip-a-coin features.

[3] In the extreme outcome condition, if a participant rolled the die more than four times, the program first presented an outcome of "2" and then presented random outcomes.

[4] In the extreme outcome condition, if the participant flipped the coin more than eight times, the program first presented an outcome of tails, and then presented random outcomes.

outcomes were extreme or standard, but participants were not paid any bonus for their wins but were merely asked to count and report them. The standard condition with no bonus for wins served as a general control condition, where participants had no incentive to misreport the number of wins, and we could test if they could accurately count them. Importantly, we also expected participants in the extreme conditions to *under*report more when they were paid for wins than they were not, even though they would receive less money. This is because they would be especially concerned with appearing dishonest if they report many wins when they are paid bonus for their wins. Such findings would support our argument that it is concern with appearing dishonest (rather than a counting problem) that drives the observed underreporting behavior.

Method.

Participants. Two hundred thirty-five undergraduate students in an Israeli university (56% females, $M_{age} = 23.91$, $SD_{age} = 3.22$) participated in this study. The data of 10 participants were not recorded due to technical errors in the program. All participants received a basic payment of 8 NIS (equivalent to about $2), and some also received a bonus as explained below. This was the first study we ran in this line of research. We conducted a post hoc power analysis using G-Power (Faul, Erdfelder, Lang, & Buchner, 2007), based on the obtained effect sizes for the interaction between outcome and bonus and for the contrast between the extreme outcome conditions (see below). The achieved power was 0.96 and 0.57, respectively. To increase the power of the contrast analysis, we replicated the extreme outcome conditions in Study 2c.

Procedure. Participants were assigned to one of four conditions, according to a 2 × 2 design, with outcome (extreme or random) and bonus (yes or no) as between-participants factors. The procedure for the extreme and standard outcome conditions with bonus was identical to that of Study 2a. The only difference in the conditions with no bonus was that participants were not promised any additional payment for their wins (both in the die roll and in the coin flip tasks). Participants in the bonus conditions were promised 0.5 NIS bonus for each win they reported, as in Study 2a.

Individual differences measures. At the end of the study, we asked participants about their general concern with appearance in everyday life. Our goal was to capture some stable differences in individual characteristics related to concern with appearance. We explored whether participants' tendency to overreport, truthfully report, or underreport their outcomes was related to these individual characteristics (i.e., whether participants who underreport their outcomes in this experiment are generally more concerned with what others think about them).

The participants were asked to answer seven questions (six of which were used in Supplemental Study S1 in the online supplemental material). First, they had to rate how important it was for them that other people would think that they were social, on a scale ranging from 1 (*not important at all*) to 7 (*very important*). The next four questions were similar, except they concerned how important it was that other people would think that they were kind, smart, honest, and unique. Next, participants had to answer three questions on a 1 to 7 scale, ranging from 1 (*not likely at all*) to 7 (*very likely*): "Imagine that a bus driver had given you too much change. How likely would you be to return the change?"; "A stranger asks you for a dollar. How likely would you be to give it to him?"; and "How likely are you to check your homework with friends, when instructed not to?" These measures included three questions that asked directly about concern with appearing honest (how important it is to appear honest, the bus driver question, and the homework question). There were also three questions that asked about concern with appearing social (how important it is to appear social, kind, and the dollar to a stranger question). Two questions (appearing smart and unique) examined general concern with appearance, not directly tied to prosocial or honesty aspects. Finally, the participants were asked to fill out some demographical questions and were debriefed.

Results.

Lying. For each participant, we computed a "lying score" by subtracting the actual number of wins (total of 5s and 6s in the die rolls and heads in the coin flips, as recorded by the program) from the number of wins she reported. To compare the magnitude of lying in the different conditions, we conducted a two-way ANOVA with outcome (extreme or random) and bonus (yes or no) as between-participants factors. The dependent measure was participants' lying score. Consistent with our hypothesis, participants' lying score was lower (i.e., more underreporting) in the extreme outcome conditions ($M = -0.47$, $SD = 0.82$) than in the random outcome conditions ($M = 0.19$, $SD = 0.87$), $F(1, 221) = 34.87$, $p < .001$, $\eta_p^2 = 0.14$. There was no main effect for bonus, $F(1, 221) = 0.35$, $p = .554$. A significant interaction occurred between outcome and bonus, $F(1, 221) = 13.64$, $p < .001$, $\eta_p^2 = 0.06$. Specifically, in the random outcome conditions, participants' lying score was higher (i.e., they overreported their outcomes more) when they gained a bonus for wins than when they did not, $F(1, 221) = 9.47$, $p = .002$, $\eta_p^2 = 0.04$. By contrast, in the extreme outcome conditions, participants' lying score was lower when they were paid a bonus for wins than when they were not, $F(1, 221) = 4.66$, $p = .032$, $\eta_p^2 = 0.02$. This suggests that in contrast to the random conditions, and in line with our prediction, in the extreme outcome conditions participants underreported their results more when they were paid a bonus for wins than when they were not. We assume that participants who reported getting 12 out of 12 wins were worried that the experimenter might think that they misreported their true outcome to get more "wins," but this concern was stronger for those who were paid for their wins, leading them to underreport more.

Next, we tested participants' lying score in each condition separately. In the random outcome conditions, when there was no bonus, participants' lying score was not different from zero ($M = -0.05$, $SD = 0.35$), $t(55) = 1.14$, $p = .261$. By contrast, when participants' outcomes were random, but they were paid for each win they got, their lying score was greater than zero, suggesting that they lied to gain more bonus ($M = 0.42$, $SD = 1.12$), $t(59) = 2.87$, $p = .006$, $d = 0.75$. In the extreme outcome conditions, when participants saw 12 wins out of 12, their lying scores were lower than zero in both the bonus ($M = -0.63$, $SD = 1.01$) and no-bonus conditions ($M = -0.30$, $SD = 0.50$), suggesting they had underreported the number of wins they got in both conditions, $t(54) = 4.69$, $p < .001$, $d = 1.28$, and $t(53) = 4.35$, $p < .001$, $d = 1.20$, respectively.

We also examined whether the rate of participants who underreported their outcome differed between conditions. We conducted a logistic regression with underreporting (yes or no)

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.

as the dependent variable. Outcome and bonus were centered, and an interaction term was computed. The analysis revealed an effect of outcome. Participants were more likely to underreport when the outcome was extreme than when it was random, 6% versus 30%, respectively, Wald = 18.54, odds ratio (*OR*) = 0.13, $p < .001$. There was no difference between bonus and no-bonus conditions, 18% versus 17%, respectively, Wald = 0.47, *OR* = 1.39, $p = .492$. There was also no interaction effect, Wald = 2.26, *OR* = 0.49, $p = .133$. In particular, in the extreme outcome condition, there was no difference between the rate of participants who underreported their outcome in the bonus condition (35%) and in the no-bonus condition (26%), $\chi^2(N = 115) = 0.91$, $p = .341$. Thus, the rate of participants who underreported their outcome in the bonus condition was not significantly higher than in the no-bonus condition, but they did so to a larger extent.

Individual differences. For each participant, we computed a "concern with appearing honest" score, by averaging the answers to the three relevant questions (after reverse-coding participants' answer to the homework question, rs = .21 and .35, ps $< .003$); we computed a "concern with appearing social" score, by averaging their answers to two of the relevant questions, $r = .60$, ps $< .001$;[5] we computed a "concern with appearing smart" score, by averaging their answers to the two relevant questions, $r = .47$, $p < .001$. In the random outcomes conditions, in both the bonus and no-bonus conditions, there were no significant differences in any of these scores between participants who *over*reported their outcomes and those who did not, ts < 1.10, ps $> .290$. In the random outcomes bonus and no-bonus conditions, there were also no significant differences in any of these scores between participants who *under*reported their outcomes and those who did not, ts < 1.59, ps $> .120$. In the extreme outcomes bonus condition, there was no significant difference in any of the scores between participants who *under*reported their outcomes and those who did not, ts < 0.93, ps $> .360$. Finally, in the extreme outcomes no-bonus condition, there was no significant difference in the "concern with appearing honest" score and the "concern with appearing smart" score, between participants who *under*reported their outcomes and those who did not, ts < 0.410, ps $> .680$. However, there was a significant difference in the "concern with appearing social" score, whereby participants who underreported their outcome had a higher concern score ($M = 5.54$, $SD = 0.69$), than participants who did not underreport their outcome ($M = 4.89$, $SD = 0.74$), $t(52) = 2.79$, $p = .007$, $d = 0.77$. Taken together, it seems that there was no clear pattern of differences in self-report of importance of appearance between participants who underreported or told the truth. Future research should further investigate individual differences related to lying to appear honest.

Discussion. The results of Study 2b provided additional evidence that participants would underreport extremely favorable outcomes, even when underreporting is costly to them. This does not seem to be a technical error due to participants making mistakes in their reporting, as participants who received random outcomes and were not paid for wins accurately reported their outcomes. Furthermore, participants who received extreme outcomes underreported their results more when they were paid for wins (i.e., a situation where the experimenter would be more likely to think that the extreme outcomes were a lie) than when they were not.

Finally, we were not able to detect any meaningful correlations between participants' likelihood to underreport and any other individual characteristics. We make no strong interpretation of this null result as it may be due to social desirability in answering hypothetical questions on honesty, tendency to help, and so forth.

Study 2c

Study 2b documented participants' tendency to underreport highly favorable outcomes that might make them appear as liars. To ensure that the study has enough power, and that participants were intentionally underreporting their outcomes, Study 2c replicated the "extreme outcome" conditions using the same behavioral paradigm (this time with an American sample).

Method.

Participants. A sample-size analysis for detecting the difference between the extreme outcome conditions, which was estimated as small in Study 2b, suggested that doubling the sample size would provide 0.9 power. We thus aimed for the same sample size as in Study 2b (extreme outcome conditions only). One hundred and five undergraduate students in an American university participated in this study in exchange for a base payment of $4. The data of four participants were not recorded due to technical errors. Because of a separate technical error, the demographic details of 29 of the participants were not recorded (but all their other data was). For the 72 participants whose demographic details we have, 56% were females, $M_{age} = 20.88$, $SD_{age} = 5.13$.

Procedure. The study followed the procedure of the extreme outcome conditions of Study 2b. Importantly, the participants were assigned to either the bonus or no-bonus conditions. In the bonus conditions, participants were paid a 25 cent bonus for each win they reported (5 or 6 on the die, or heads with the coin). In the no-bonus conditions, participants were not promised any bonus.

Results.

Lying. Participants' lying scores in both the bonus and no-bonus conditions were significantly lower than zero, confirming our main hypothesis that participants would underreport their wins, $t(51) = 4.97$, $p < .001$, $d = 0.69$, and $t(48) = 3.51$, $p = .001$, $d = 0.56$, respectively. In addition, participants underreported the number of wins they received more when they got a bonus for wins ($M = -0.52$, $SD = 0.75$) than when they did not ($M = -0.20$, $SD = 0.41$), $t(99) = 2.59$, $p = .011$, $d = 0.53$. Participants were also more likely to underreport in the bonus condition (37%) than in the no-bonus condition (20%), $\chi^2(N = 101) = 3.21$, $p = .047$. Thus, not only were there differences in the magnitude of underreporting between the conditions, but also in the number of participants underreporting.

Discussion. Study 2c replicated our main finding from Study 2b that participants underreported the number of wins in a string of wins and did so more strongly when they were paid for wins than when they were not paid for wins. This is in line with our prediction that when people are paid a bonus for their wins, they will be more concerned with appearing dishonest when reporting

[5] We did not add the question about the driver, because it did not correlate strongly with any of the questions. The correlations with the social questions, with which we expected it to correlate most, were $r = 0.04$, $p = .53$ with the social question, and $r = 0.14$, $p = .04$ with the kind question.

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.

Table 1
Rates of Underreporting in Studies 1–2 by Study and Condition

Study	% underreporting in standard outcome condition	% underreporting in extreme outcome condition	*n*	Statistic	Effect size φ
Study 1a	0%	18%	115	Fisher's test $p < .001^{***}$	.31
Study 1b	0%	8%	156	Fisher's test $p < .05^{*}$	.20
Study 1c	5%	12%	184	$\chi^2 = 4.23$	.13
Study 1d (overreporting not revealed)	0%	6%	245	Fisher's test $p < .05^{*}$	.01
Study 2a	5%	24%	137	$\chi^2 = 10.32^{***}$	.28
Study 2b (bonus)	3%	35%	115	Fisher's test $p < .001^{***}$	.13
Study 2b (no bonus)	9%	26%	110	Fisher's test $p < .05^{*}$	.23

* $p \leq .05$. *** $p \leq .001$.

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.

an extreme outcome. See Table 1 for summary of findings from Studies 1–2.

Study 3

Studies 1a–1d and 2a–2c provided support for our claim that people will sometimes underreport their favorable outcomes to avoid appearing dishonest. In Studies 3a–3b, we explored how people judge others who underreport or tell the truth about receiving extreme outcomes. We use these data to examine whether people's concern with appearing dishonest in such cases is actually valid.

Study 3a

In Study 3a, we tested whether observers would perceive a person who reported an extreme outcome (i.e., eight heads out of eight in Studies 2a and 2b) as less moral or honest than a person who reported a standard outcome (i.e., four heads out of eight). If observers negatively evaluate people for reporting very favorable outcomes, then it would provide some evidence that people's concerns with appearing dishonest are valid (i.e., that they will actually be judged negatively for reporting such extreme outcomes).

Method.

Participants. Participants were recruited online using the Amazon Mechanical Turk platform; participation was restricted to adult participants from the United States. One hundred fifty adults (53% females, $M_{age} = 35.56$, $SD_{age} = 10.41$) participated in this 5-min study for 25 cents. We preregistered our hypothesis and sample size on https://aspredicted.org/pr4z9.pdf. According to a preregistered rule, we excluded from further analyses all participants who failed to answer correctly the reading comprehension question. Eight participants were excluded based on this rule.

Procedure. The participants read the following scenario:

> We have recently conducted an experiment in our lab where participants were asked to flip a coin eight times, and report to us the number of heads they received. Each participant flipped the coin in private. We ran 500 participants in total. Participants knew that on top of the flat fee that they would be paid for this study, they would earn a bonus of 25 cents for each head they reported.

Participants were then randomly assigned into one of two conditions—extreme outcome condition or standard outcome condition. Participants in the extreme outcome condition were asked to imagine a participant named Jim who reported flipping eight heads (out of eight). Participants in the standard outcome condition were instead asked to imagine a participant named Jim who reported flipping four heads (out of eight). All participants were asked to answer five questions on the morality of Jim. These questions were based on questions used in previous research for judging morality of liars (Levine & Schweitzer, 2014). Specifically, participants were asked to rate on a scale ranging from 1 (*not at all*) to 7 (*extremely*) how ethical they thought Jim was, given the outcome he reported, how moral he was, and how good he was. They were also asked to rate on a scale ranging from 1 (*strongly disagree*) to 7 (*strongly agree*) whether Jim was honest and whether he was telling the truth. Participants were then asked one comprehension question ("How many times was Jim asked to flip the coin?"), and a few demographic questions.

Results. We computed for each participant a "morality score" by averaging the answers to the five questions on the morality of Jim (*rs* ranged from .73 to .89, *ps* $< .001$). As expected, participants found Jim to be less moral when he reported getting eight out of eight heads ($M = 3.00$, $SD = 1.54$) than when he reported getting four out of eight heads ($M = 4.99$, $SD = 1.16$), $t(140) = 8.68$, $p < .001$, $d = 1.47$. This suggests that participants indeed judged others more harshly when they reported outcomes that seemed "too good to be true."

Discussion. The results of Study 3a demonstrated that observers perceived people who reported outcomes that were extremely favorable to the self as less moral and honest than those who reported standard outcomes. Clearly, extremely favorable outcomes sometimes occur (in the scenario described here, there was about 2% chance that one of the 500 participants will flip eight out of eight heads). Yet, when observers judge someone who reports such an extreme outcome, they tend to attribute the outcome to this person's dishonesty rather than to extreme luck. If people do actually judge others negatively for reporting such outcomes, it is plausible that when people consider how to report their own extreme outcomes, they consider the fact that others will judge them negatively and take precautions by underreporting the truth (indeed, findings from our previous studies reveal that these concerns with being perceived as dishonest do in fact mediate people's choice to underreport).

Study 3b

In Study 3a, we found that observers judged a person who reported an extreme outcome more negatively than a person who

reported a standard outcome. In Study 3b, we used a similar paradigm to that used in Study 3a, adding a condition where a person reported a highly favorable outcome, but not the most extreme one (i.e., flipping seven heads out of eight flips). We predicted that observers would judge this person more negatively than a person who reported a standard outcome (four out of eight), but more positively than a person who reported the most extreme outcome (eight out of eight), because the outcome is more likely to be true. This would mean that people's motivation to lie to appear honest is at least partially justified.

In Study 3b, we also investigated what observers thought of people who they knew were telling the truth, or underreporting it. To test this, we added three conditions where the protagonist reported the same outcome (four, seven, or eight heads out of eight), but the participants also knew what the true outcome was (four, eight, or eight heads out of eight). We hypothesized that a protagonist who lied to appear honest would be judged as less honest than a protagonist who reported true outcomes (because there would be evidence that s/he lied), but that this would not lead observers to the conclusion that the liar was a bad person (because s/he was lying against his or her self-interest).

Method.

Participants. Participants were recruited online using the Amazon Mechanical Turk platform; participation was restricted to adult participants from the United States. Three hundred adults (47% females, $M_{age} = 34.40$, $SD_{age} = 11.34$) participated in this 5-min study for 25 cents. We preregistered our hypothesis and sample size on https://aspredicted.org/nb5gu.pdf. According to a preregistered rule, we excluded from further analyses all participants who failed to answer correctly at least one of the two reading comprehension questions. Forty participants were excluded based on this rule.

Procedure. Each participant read three short scenarios that were similar to the scenario used in Study 3a. Each of the three scenarios depicted a protagonist who flipped a coin in private eight times as part of an experiment in the lab and was compensated according the outcome he reported (earning money for each time he flipped heads). Specifically, Jim reported flipping eight heads out of eight, Dan reported flipping four heads out of eight, and Mark reported flipping seven heads out of eight. The presentation order of the three protagonists was counterbalanced across participants.

Participants were randomly assigned into one of two between participants conditions—truth not-revealed condition or truth revealed condition. In the truth not-revealed condition, just as in the scenario in Study 3a, participants found out only about the outcome reported by the protagonist (e.g., "Imagine a participant named Mark who reported flipping four heads out of eight"). In the truth-revealed condition, participants found out about the protagonist's true outcome, and also about the outcome he reported (e.g., "Imagine a participant named Mark who flipped four heads out of eight. Mark then reported flipping four heads out of eight.") Importantly, in the truth-revealed condition, participants learned that Mark truthfully reported his standard outcome (four heads out of eight), Jim truthfully reported his extreme outcome (eight heads out of eight), and Dan underreported his extreme outcome (reporting seven heads when he truly flipped eight heads out of eight). This design allowed us to compare observers' judgment of a protagonist who had underreported his extreme outcome and their judgment of a protagonist who had truthfully reported his standard or extreme outcomes.

After each of the three scenarios, participants were asked (a) how much they thought the person they read about was a good person, and (b) how honest this person was, given this person's reported outcome, and his true outcome—when this was revealed. Participants indicated their ratings on scales ranging from 1 (*not at all*) to 7 (*extremely*). Participants were asked an open-ended comprehension question ("How many times was each participant in the study asked to flip the coin?"), and a multiple-choice question ("Did participants who reported more 'heads' get more bonus?" They could choose between: "yes," "no," and "I do not know"). They then completed a few demographic questions.

This 3 (outcome: standard, extreme, less-than-extreme) $\times$ 2 (truth revealed: yes or no) mixed design allowed us to investigate how observers (who evaluate different reports sequentially) judge people who report different outcomes when their true outcomes are known or not. When the truth is not revealed, we expected observers to judge the protagonist more harshly the more extreme the outcome he reports (in line with our findings from Study 3a). When observers find that the protagonist is telling the truth (or underreporting it), we expected their judgment to be more positive, particularly for the more extreme outcomes. Importantly, we expected observers to judge the protagonist who lied and underreported his extreme outcome as dishonest, but as a good person—just as the one who truthfully reported his outcome.

Results. We conducted a 2 $\times$ 3 mixed design ANOVA with a between-participants factor (truth revealed or not) and a within-participants factor (reported outcome: four, seven, or eight). For means and standard deviations by condition, see Table 2. We ran two analyses, one where the dependent variable was participants' ratings of the protagonist as a good person, and a second where the dependent variable was their rating of his honesty. Extending the findings from Study 3a, we found that the outcome reported by the protagonist affected the participants' judgment of how good a person he was, $F(2, 257) = 59.80$, $p < .001$, $\eta_p^2 = 0.32$ and participants' judgment of the protagonist's honesty, $F(2, 257) = 194.15$, $p < .001$, $\eta_p^2 = 0.60$. Further, consistent with our hypothesis, participants found the protagonist to be a better person when his true outcome was revealed than when it was not, $F(1, 258) = 23.64$, $p < .001$, $\eta_p^2 = 0.08$. Participants also found the protagonist to be more honest when his true outcome was revealed than not, $F(1, 258) = 34.37$, $p < .001$, $\eta_p^2 = 0.12$. There was no interaction between the reported number and the truth factor in participants' judgment of how good a person the protagonist was, $F(2, 257) = 1.35$, $p = .261$. However, there was a significant interaction between the reported number and the truth factor in participants judgment of the protagonist's honesty, $F(2, 257) = 21.36$, $p < .001$, $\eta_p^2 = 0.14$.

Next, we ran planned contrasts, according to the preregistration. We found that participants believed the protagonist to be a better person when they knew that the outcome he reported was true (than when they did not know), both when the reported outcome was standard (four out of eight heads), $t(258) = 3.79$, $p < .001$, $d = 0.47$, and extreme (eight out of eight heads), $t(258) = 4.01$, $p < .001$, $d = 0.50$. Participants also believed the protagonist to be a better person when they knew that he had underreported an extreme outcome (seven out of eight heads), than when he reported seven out of eight heads but they did not know what the true

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.

Table 2
Participants' Average Ratings of How Good and How Honest the Protagonist Is, Depending on the Protagonist's Reported Outcome and Whether or Not the Protagonist's True Outcome Was Revealed to Them, in Study 3b

Reported outcome [True outcome]	4 [4]	7 [8]	8 [8]
Good person			
True outcome not revealed	5.04 (1.25)	3.99 (1.42)	3.56 (1.56)
True outcome revealed	5.59 (1.04)	4.62 (1.60)	4.43 (1.97)
Honest			
True outcome not revealed	5.46 (1.20)	3.25 (1.59)	2.65 (1.73)
True outcome revealed	6.01 (1.16)	3.60 (1.65)	4.43 (2.34)

Note. Ratings ranged from 1 (*not at all*) to 7 (*extremely*). Standard deviations are in parentheses. Participants knew the true outcome only in the "True outcome revealed" condition.

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.

outcome was, $t(258) = 3.37$, $p = .001$, $d = 0.42$. Participants also believed the protagonist was more honest when they knew that the outcome he reported was true (than when they did not know), both when the outcome was standard, $t(258) = 3.75$, $p < .001$, $d = 0.47$, and extreme, $t(258) = 7.07$, $p < .001$, $d = 0.88$. However, they did not believe a person who underreported an extreme outcome was more honest than a person who reported an extreme outcome (and whose true outcome was not revealed), $t(258) = 1.65$, $p = .101$.

In addition, and in line with our predictions, paired *t* tests showed that when the truth was revealed and the protagonist underreported an extreme outcome, participants judged him to be just as good a person as a protagonist who truthfully reported the extreme outcome, $t(121) = 1.28$, $p = .202$. At the same time, participants judged the protagonist who underreported his outcome to be less honest, $t(121) = 3.83$, $p < .001$, $d = 0.11$. Thus, it seems that although participants realized that underreporting was a lie, it did not take a significant toll on the moral image of the protagonist in their eyes. Note that unlike our prediction, a protagonist who truthfully reported a standard outcome was judged as a better person than a protagonist who underreported an extreme outcome, $t(121) = 6.83$, $p < .001$, $d = 0.72$. However, this difference seems to be driven by the report of an extreme outcome, rather than by the underreporting, as a protagonist who truthfully reported a standard outcome was also judged as a better person than a protagonist who truthfully reported an extreme outcome, $t(121) = 7.17$, $p < .001$, $d = 0.74$. Participants may have attributed extreme outcomes—even when told they were true—to the protagonists' actions.

Discussion. Study 3b demonstrates that when observers do not know a person's true outcome, underreporting an extreme outcome (by a little) helps in preserving a more moral and honest appearance. Thus, when the outcomes are not verified to be true, participants who lie to appear honest do appear more honest than those who report an extreme true outcome.

Further, this study also demonstrates that when observers happen to find out that a person has underreported an extreme outcome, they find it to be dishonest, but do not find the person to be a worse person (compared to a person who truthfully reported the same extreme outcome). Thus, underreporting has some advantage when the truth is unknown, and it does not make one look like a bad person when the truth is uncovered.

General Discussion

Our findings highlight an understudied motivation for dishonesty: lying to appear honest. Past research demonstrates that people care about having a reputation as an honest person (Batson et al., 1997; Gneezy et al., 2018). Whereas this desire to appear honest commonly leads people to lie *less* (Gneezy et al., 2018), we show that in some cases, it may actually lead people to lie *more*. In particular, we have demonstrated that some people who receive extremely favorable outcomes in private will report in public that they received less favorable outcomes. For example, participants in Study 1c, who played the role of employees and happened to drive the maximal reimbursable mileage, reported driving fewer miles, thereby forgoing some of the reimbursement they deserved. Participants who played a chance game in the lab in Studies 2a–2c and won in private in each and every round reported less favorable outcomes than they actually achieved. Participants intentionally misrepresented their true outcomes, thereby breaking the moral principle of telling the truth and giving up some payment they deserved.

We propose that what motivated participants to lie in our experiments is their concern that if they tell the truth, others may think they are lying. In other words, participants overcame their aversion toward lying and the monetary costs involved in lying in order to appear honest to someone else. Our studies provide support for this account. In Studies 1b, 1c, and 2a (see also S1 in the online supplemental material), participants' judgment of the probability that others would perceive them as liars accounted (at least partially) for their likelihood to underreport their outcomes. For example, participants in the role of workers in Study 1c, who drove the maximal number of miles for which they could be reimbursed, thought it was more likely that their manager would suspect they lied, compared to participants who drove a standard number of miles. Accordingly, the participants who drove an extreme number of miles underreported their mileage more. In Study 1d, the probability that the manager would assume that an extreme report is dishonest moderated participants' likelihood of underreporting their mileage. More support for this account comes from our behavioral studies (2a–2c). Specifically, we measured the number of wins that participants reported upon rolling a die and flipping a coin. We found participants who won every single die

roll and coin toss underreported their wins and that participants underreported their number of wins more when they received a bonus for their wins than when they did not receive any bonus. We argue that they acted counter to their financial interests (i.e., reporting fewer wins when wins were worth money) because having financial interests made them especially vulnerable to be seen by others as selfish liars. Thus, when the likelihood of being seen as dishonest when telling the truth is higher, people are more likely to lie to appear honest. Indeed, concern with appearing dishonest in the eyes of the experimenter mediated participants' degree of underreporting (Study 2a). We note also that it is possible that had the incentives been even greater, people would have reported the truth to capitalize on the incentives. Clearly, people will need to balance their concerns with appearing dishonest against their desire to maximize resources, and so people would likely tell the truth if the monetary incentives were strong enough.

Our findings suggest that when people obtain extremely favorable outcomes, they anticipate other people's suspicious reactions and prefer lying and appearing honest over telling the truth and appearing as selfish liars. Is people's concern with being judged as dishonest when reporting extremely favorable outcomes (which cannot be verified) valid? The results of Studies 3a and 3b suggest that it is. In these studies, we found that the more extreme the outcomes a person reported, the more dishonest and immoral she appeared. Thus, honestly reporting extremely favorable outcomes may indeed have a negative impact on one's honest reputation.

We argue that lying to appear honest should be more prevalent and likely when one is particularly concerned with appearing dishonest. Thus, when a person expects others to believe her even when she reports statistically unlikely outcomes (e.g., because she has a longstanding honest reputation), or if she does not care whether others believe her or not, we do not expect to see any lying. When concern with appearing dishonest is great, however, we expect some people to lie to appear honest. Such lying can take the form of underreporting, but also the form of overreporting, depending on what direction might help preserve an honest appearance more. We speculate, for example, that following the Open Science revolution that put special emphasis on honesty and transparency in research (Nosek et al., 2015; Simmons, Nelson, & Simonsohn, 2011), researchers became more concerned also about true outcomes that may seem to reviewers "too good to be true." For instance, researchers who in the past may have felt the urge to round marginal results *down* to $p < .05$ may now feel the urge to round them *up* to $p > .05$, so that the results do not appear to have been *p*-hacked. In a different vein, an Amazon customer who complained about a missing package last week and received a refund may be reluctant to complain again this week when an unrelated package also goes missing. This is because the customer may worry that Amazon is monitoring the complaints and will suspect she is trying to cheat and gain from overreporting missing packages. In these real-life examples, just like in our studies, concern with appearing dishonest may erroneously lead people to misreport their outcomes.

This research joins recent work showing how some types of lying are not so negative. Research has found that people judge the morality of lies depending on the motivation of the liar. Specifically, people may judge others who tell prosocial lies (lies that benefit others) as more moral than those who tell a truth that is less favorable to others (Levine & Schweitzer, 2014). Thus, when judging the morality of one's action, judges may give more weight to its prosocial nature than to its dishonesty. Relatedly, lying in which one falsely gives someone else credit for one's own idea is seen much less negatively than when one falsely takes credit for someone else's idea (Silver & Shaw, 2018). Future research should examine how the motivation to appear honest affects moral judgment. We predict that lies intended to preserve one's honest reputation will be evaluated as less immoral than lies intended to attain a monetary benefit (even when both lies diverge from the truth to the same extent). The reason is that those who lie to appear honest, unlike the "typical" liars, are not motivated by greed. At the same time, we predict that lying to appear honest will not be perceived as favorably as prosocial lying because its ultimate goal is, after all, selfish (i.e., maintaining one's favorable reputation).

The paradigm developed in our studies allows us to easily examine lying at the individual level (i.e., we can directly tell how much each participant lied). In our studies (Study 2b and S1 in the online supplemental material), we were not able to correlate participants' lying behavior with self-reported individual differences. Still, our paradigm allows for such investigation. Previous research on lying has struggled with measuring lies at the individual level, while maintaining the confidentiality of participants' decisions in the experiment. Studies of unethical behavior have often measured lying at an aggregated level by comparing the average report across a group of individuals to the expected outcome (of a random device, see Shalvi et al., 2011; or of a control group who could not lie, see Mazar et al., 2008). Some researchers who investigated individual differences in lying relied on statistical models and simulations to differentiate between individuals who lied and those who told the truth (Heck, Thielmann, Moshagen, & Hilbig, 2018; Moshagen & Hilbig, 2017). Others have designed other methods for detecting lies at the individual level (e.g., Gneezy, Rochenbach, & Serra-Garcia, 2013). Our new method provides researchers with a useful way to identify liars at the individual level and correlate their behavior with personality traits or other individual measures. This would be useful for understanding the individual characteristics of people who are willing to lie to gain money (or to help others).

Another interesting future direction would be to identify the individual characteristics of people who lie to appear honest. Indeed, in our studies, only a minority of participants lied to appear honest, and it would be informative to understand what characterizes them and possibly differentiates them from those who lie to attain a material reward and those who lie to help others. We suspect that these participants are generally more sensitive to their appearance in the eyes of others.

Although we focus on cases where dishonest underreporting is aimed at maintaining an honest reputation, we also acknowledge there are of course other reasons for dishonest underreporting. First, people may underreport their achievements or conceal their higher status out of modesty, or to avoid negative reactions by others to their exceptional achievements (e.g., envy; Arnett & Sidanius, 2018; Phillips, Rothbard, & Dumas, 2009). Relatedly, sandbagging is common in sporting contexts and involves a person dishonestly pretending to be less good than she actually is to convince her opponent to try less hard or to entice betting action (Gibson & Sachau, 2000). People may also be dishonest about truths that appear beneficial to them for other strategic reasons: A lawyer may underbill a customer now to attract more new business

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.

later, and an innocent suspect may plead guilty to escape a harsh interrogation (Kassin, 2015). In such cases, people lie to secure some direct material gain. In this article, we have focused on cases where people do not secure direct material gains (or avoid direct costs) by underreporting their outcomes—in fact, they forgo immediate benefits. Still, we documented clear lying behavior. Therefore, in our experiments, the only benefit in lying is maintaining an honest reputation. Whereas in our experiments people cared about an honest appearance in and of itself, in many other situations lying can, of course, provide benefits down the road.

Final Remark

Although Akerlof's (1983) conclusion that being honest is the best way to appear honest may hold true in most situations, our results suggest a modest revision. In situations where being honest seems suspicious, people who want to appear honest may be better off lying. Our findings suggest that people are quick to realize this and, accordingly, sometimes lie to appear honest.

Context Paragraph

For several years we have been studying how people are motivated by how they appear to others and, in particular, how they must balance different appearances. For example, we have found that people will sometimes forgo generosity to others in order to avoid *appearing* biased or unfair (Choshen-Hillel, Shaw, & Caruso, 2015; Shaw et al., 2018). In the current research, we have extended our research to explore how people's concern with appearance may affect their behavior in the domain of unethical behavior and lying. Specifically, we explored cases in which people's desire to appear honest to others may, ironically, lead them to lie to others. Namely, we looked at situations where the truth appeared too good to be true. Here, in different samples of attorneys, scenario studies, and consequential behavioral paradigms, we found robust support for our claim that people will lie to appear honest.

References

Akerlof, G. A. (1983). Loyalty filters. *The American Economic Review, 73,* 54–63.

Andreoni, J., & Bernheim, B. (2009). Social image and the 50–50 norm: A theoretical and experimental analysis of audience effects. *Econometrica, 77,* 1607–1636. http://dx.doi.org/10.3982/ECTA7384

Arnett, R. D., & Sidanius, J. (2018). Sacrificing status for social harmony: Concealing relatively high status identities for one's peers. *Organizational Behavior and Human Decision Processes, 147,* 108–126. http://dx.doi.org/10.1016/j.obhdp.2018.05.009

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

Bryan, C. J., Adams, G. S., & Monin, B. (2013). When cheating would make you a cheater: Implicating the self prevents unethical behavior. *Journal of Experimental Psychology: General, 142,* 1001–1005. http://dx.doi.org/10.1037/a0030655

Caruso, E. M., & Gino, F. (2011). Blind ethics: Closing one's eyes polarizes moral judgments and discourages dishonest behavior. *Cognition, 118,* 280–285. http://dx.doi.org/10.1016/j.cognition.2010.11.008

Choshen-Hillel, S., Shaw, A., & Caruso, E. M. (2015). Waste management: How reducing partiality can promote efficient resource allocation. *Journal of Personality and Social Psychology, 119,* 210–231.

Cohn, A., Fehr, E., & Maréchal, M. A. (2014). Business culture and dishonesty in the banking industry. *Nature, 516,* 86–89. http://dx.doi.org/10.1038/nature13977

Dana, J., Cain, D. M., & Dawes, R. M. (2006). What you don't know won't hurt me: Costly (but quiet) exit in dictator games. *Organizational Behavior and Human Decision Processes, 100,* 193–201. http://dx.doi.org/10.1016/j.obhdp.2005.10.001

DePaulo, B. M., Kashy, D. A., Kirkendol, S. E., Wyer, M. M., & Epstein, J. A. (1996). Lying in everyday life. *Journal of Personality and Social Psychology, 70,* 979–995. http://dx.doi.org/10.1037/0022-3514.70.5.979

Erat, S., & Gneezy, U. (2012). White lies. *Management Science, 58,* 723–733. http://dx.doi.org/10.1287/mnsc.1110.1449

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,* 175–191. http://dx.doi.org/10.3758/BF03193146

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

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

Gibson, B., & Sachau, D. A. (2000). Sandbagging as a self-presentational strategy: Claiming to be less than you are. *Personality and Social Psychology Bulletin, 26,* 56–70. http://dx.doi.org/10.1177/0146167200261006

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., 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

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

Gneezy, U., Rochenbach, B., & Serra-Garcia, M. (2013). Measuring lying aversion. *Journal of Economic Behavior & Organization, 93,* 293–300. http://dx.doi.org/10.1016/j.jebo.2013.03.025

Grice, H. P. (1991). Logic and conversation. In S. Davis (Ed.), *Pragmatics: A reader* (pp. 305–315). New York, NY: Oxford University Press. http://dx.doi.org/10.1163/9789004368811_003

Hayes, A. F. (2013). *Introduction to mediation, moderation, and conditional process analysis: A regression-based approach*. New York, NY: Guilford Press.

Heck, D. W., Thielmann, I., Moshagen, M., & Hilbig, B. E. (2018). Who lies? A large-scale reanalysis linking basic personality traits to unethical decision making. *Judgment and Decision Making, 13,* 356–371.

Hofmann, W., Wisneski, D. C., Brandt, M. J., & Skitka, L. J. (2014). Morality in everyday life. *Science, 345,* 1340–1343.

Kassin, S. M. (2015). The social psychology of false confessions. *Social Issues and Policy Review, 9,* 25–51. http://dx.doi.org/10.1111/sipr.12009

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.

Levine, E., Hart, J., Moore, K., Rubin, E., Yadav, K., & Halpern, S. (2018). The surprising costs of silence: Asymmetric preferences for prosocial lies of commission and omission. *Journal of Personality and Social Psychology, 114,* 29–51. http://dx.doi.org/10.1037/pspa0000101

Levine, E. E., & Schweitzer, M. E. (2014). Are liars ethical? On the tension between benevolence and honesty. *Journal of Experimental Social Psychology, 53,* 107–117. http://dx.doi.org/10.1016/j.jesp.2014.03.005

Levine, E. E., & Schweitzer, M. E. (2015). Prosocial lies: When deception breeds trust. *Organizational Behavior and Human Decision Processes, 126,* 88–106. http://dx.doi.org/10.1016/j.obhdp.2014.10.007

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

Mazur, M. J., & Plumley, A. H. (2007). Understanding the tax gap. *National Tax Journal, 60,* 569–576. http://dx.doi.org/10.17310/ntj.2007.3.14

Moshagen, M., & Hilbig, B. E. (2017). The statistical analysis of cheating paradigms. *Behavior Research Methods, 49,* 724–732. http://dx.doi.org/10.3758/s13428-016-0729-x

Nosek, B. A., Alter, G., Banks, G. C., Borsboom, D., Bowman, S. D., Breckler, S. J., . . . Yarkoni, T. (2015). Promoting an open research culture. *Science, 348,* 1422–1425. http://dx.doi.org/10.1126/science.aab2374

Phillips, K. W., Rothbard, N. P., & Dumas, T. L. (2009). To disclose or not to disclose? Status distance and self-disclosure in diverse environments. *The Academy of Management Review, 34,* 710–732.

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

Schurr, A., Ritov, I., Kareev, Y., & Avrahami, J. (2012). Is that the answer you had in mind? The effect of perspective on unethical behavior. *Judgment and Decision Making, 7,* 679–688.

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., Gino, F., Barkan, R., & Ayal, S. (2015). Self-serving justifications: Doing wrong and feeling moral. *Current Directions in Psychological Science, 24,* 125–130. http://dx.doi.org/10.1177/0963721414553264

Shaw, A., Choshen-Hillel, S., & Caruso, E. M. (2018). Being biased against friends to appear unbiased. *Journal of Experimental Social Psychology, 78,* 104–115. http://dx.doi.org/10.1016/j.jesp.2018.05.009

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.

Silver, I., & Shaw, A. (2018). No harm, still foul: Concerns about reputation drive dislike of harmless plagiarizers. *Cognitive Science, 42*(Suppl. 1), 213–240. http://dx.doi.org/10.1111/cogs.12500

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,* 1359–1366. http://dx.doi.org/10.1177/0956797611417632

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

Weisel, O., & Shalvi, S. (2015). The collaborative roots of corruption. *Proceedings of the National Academy of Sciences of the United States of America, 112,* 10651–10656. http://dx.doi.org/10.1073/pnas.1423035112

Received February 6, 2019
Revision received December 2, 2019
Accepted December 17, 2019 ■

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.