Lying is sometimes ethical, but honesty is the best policy: The desire to avoid harmful lies leads to moral preferences for unconditional honesty

Sarah Jensen*[a], Emma E. Levine*[b], Michael W. White*[b], Elizabeth Huppert[d]

*the first three authors are listed alphabetically and contributed equally

[a] Eccles School of Business, University of Utah
[b] The University of Chicago Booth School of Business
[c] Columbia Business School, Columbia University
[d] Kellogg School of Management, Northwestern University

Forthcoming at the Journal of Experimental Psychology: General

Address correspondence to Emma E. Levine, Associate Professor of Behavioral Science, The University of Chicago Booth School of Business. Email: Emma.Levine@chicagobooth.edu, Phone: 773-834-2861. We are grateful for feedback from Dan Bartels, Jonathan Berman, Berkeley Dietvorst, Nick Epley, Celia Gaertig, Jesse Graham, Nicholas Herzog, Justin Landy, Abigail Sussman, members of the HOPE lab at the University of Chicago and attendees of the Booth Behavioral Science faculty brown bag. This paper also benefited from feedback following presentations at the *Society of Judgment and Decision-making* Annual Conference (November 2019) and the *Academy of Management* Annual Meeting (August 2020 and 2021). We are grateful for research assistance from Solomon Lister and Jordyn Schor. This research was supported by the Charles E. Merrill Faculty Research Award at the University of Chicago Booth School of Business, awarded to Emma Levine. Chicago Booth's Center for Decision Research labs helped with data collection. An earlier version of this manuscript was posted on the preprint server PsyArXiv under the title, "Lying is ethical, but honesty is the best policy: Introducing and testing a theory of moral error avoidance." All data, syntax, and materials are available at: https://tinyurl.com/HWOL-OSF.

ABSTRACT

People believe that some lies are ethical, while also claiming that "honesty is the best policy." In this article, we introduce a theory to explain this apparent inconsistency. Even though people view prosocial lies as ethical, they believe it is more important – and more moral – to avoid harmful lies than to allow prosocial lies. Unconditional honesty (simply telling the truth, without finding out how honesty will affect others) is therefore seen as ethical because it prevents the most unethical actions (i.e., harmful lies) from occurring, even though it does not optimize every moral decision. We test this theory across five focal experiments and ten supplemental studies. Consistent with our account, we find that communicators who tell the truth without finding out how honesty will affect others are viewed as more ethical, and are trusted more, than communicators who look for information about the social consequences of honesty before communicating. However, the moral preference for unconditional honesty attenuates when it is certain that looking for more information will not lead to harmful lies. Overall, this research provides a holistic understanding of how people think about honesty and suggests that moral rules are not valued because people believe all rule-violations are wrong, but rather, because they believe some violations must be avoided entirely.

Abstract word count: 211

PUBLIC SIGNIFICANCE STATEMENT

This research explains why people value unconditional honesty. Even though most people believe that some lies – namely, prosocial lies – are ethical, they believe that the moral costs of harmful lies outweigh the moral benefits of prosocial lies. A policy of unconditional honesty prohibits harmful lying. Consequently, communicators who are unconditionally honest are seen as more moral than communicators who seek out information about how honesty affects others.

Keywords: honesty, moral judgment, uncertainty, decision-making, information avoidance

Introduction

Honesty is a fundamental moral value. Its importance, emphasized across cultures and religions, is central to both moral identity and interpersonal judgment (Goodwin et al., 2014; Hartley et al., 2016). In everyday life, decisions involving honesty and dishonesty are among the most common moral decisions we make. People list honesty as the value that is most important to them (Graham et al., 2015), and honesty is the second most common moral act people encounter in their everyday lives (Hofmann et al., 2014).

Despite honesty's prominence in moral judgment and everyday decision-making, the psychology of honesty remains elusive. Even though most people lie frequently (DePaulo et al., 1996) and privately hold quite nuanced views of honesty (Levine, 2022), they tend to take absolute stances in public (e.g., claiming that lying is *never* ethical) and reward others who do the same (Huppert et al., 2023). The inconsistency between how people talk about honesty publicly and how they engage in it privately presents a challenge for understanding the moral judgment of honesty.

The moral judgment of honesty is also misunderstood because most empirical research has examined a relatively narrow range of honest and dishonest behaviors. A large body of research in behavioral ethics, organizational behavior, and economics has explored when and why people cheat, steal, and lie for personal gain. In this work, dishonesty is typically confounded with selfishness, and honesty with prosociality (e.g., Gino & Galinsky, 2012; Lee et al., 2019; Mazar et al., 2008; for reviews see: Abeler, Nosenzo, & Raymond, 2019; Gerlach, Teodorescu, & Hertwig, 2019). This work provides enormous insight into the causes of selfish dishonesty and the destructive consequences thereof. However, it provides little insight into how people think about honesty itself and how people navigate more complex dilemmas between

honesty and other moral values. An emerging body of research on morally motivated lies (e.g., Galak & Critcher, 2022; Hildreth et al., 2016; Levine & Schweitzer, 2015; Weisel & Shalvi, 2015) has begun to address this gap, but it is still limited in its focus on single acts of (dis)honesty. Recent work in moral psychology has been similarly focused on how people make sense of specific moral acts, largely concluding that the judgment of moral acts boils down to concerns about harm (e.g., Gray et al., 2014; Gray et al., 2022; Schein & Gray, 2018). People have a propensity to justify and engage in specific acts of deception, particularly when they perceive these acts as preventing harm (Levine, 2022), but this propensity is difficult to reconcile with the degree to which people endorse honesty as a policy (i.e., as a value that should be followed across time and circumstances).

In the present article, we develop and test a novel theory that explains the apparent inconsistency between moral judgments of honesty as an absolute policy and moral judgments of specific honest – or dishonest – acts. Though people believe it is sometimes ethical to tell prosocial (i.e., helpful) lies, we propose that people endorse unconditional honesty as a policy because they believe that avoiding harmful lies is more important than allowing prosocial lies. As a result, communicators who engage in unconditional honesty (by telling the truth without seeking out information about how truth-telling affects others) are seen as more ethical and are more likely to be trusted than communicators who engage in conditional honesty (by looking for information about how truth-telling affects others before making communication decisions). Taken together, these studies explain why people value absolute moral rules and those who uphold them (e.g., Jordan, Hoffman, Nowak, & Rand, 2016; Van Zant & Moore, 2015; Zlatev, 2019), despite also endorsing moral nuance. Absolute honesty is valued because people want to minimize harmful lying, not because people actually believe that all deception is wrong. These

insights add to the fundamental understanding of honesty, shed light on preferences for categorical moral rules broadly, and suggest promising avenues for future research in moral psychology.

Moral Judgments of Prosocial lies

Though most existing research on honesty and deception examines the conflict between prosocial truths and harmful lies, an emerging body of research has begun to examine the antecedents and consequences of *prosocial lies*, lies that benefit others (for review, see Levine & Lupoli, 2021). The ability to tell prosocial lies begins in childhood, with many children telling prosocial lies by age three (Talwar et al., 2007). These lies are often socially rewarded, leading them to persist into adulthood. For example, employees inflate their own and others' performances to help their work teams (Hildreth & Anderson, 2018; Weisel & Shalvi, 2015; Wiltermuth, 2011), doctors offer false hope to spare patients emotional distress at the end of their lives (Hart, 2022; Levine et al., 2018), and parents, teachers, and managers offer false praise to motivate others and avoid emotional harm (Jampol & Zayas, 2021; Lupoli, Jampol, & Oveis, 2017).

Importantly, prosocial lies, unlike harmful lies, are often seen as acceptable. For example, Levine and Schweitzer (2014) find that communicators who lie are judged to be more ethical than those who tell the truth when lying is associated with a monetary gain for a partner (and by comparison, truth is associated with a monetary cost). Recent work has also examined how people judge the ethicality of prosocial lies in everyday life, finding that most people – including communicators, targets, and third parties – believe deception is ethical when it prevents unnecessary harm to the target (Levine, 2022). Judgments of unnecessary harm hinge on two key factors: the degree to which truth-telling causes harm (i.e., emotional pain and suffering) at the

moment of communication and the degree to which the truth has instrumental value (i.e., leads to enlightenment, growth, or behavioral change). When the truth causes immediate harm and has low instrumental value, deception is perceived to prevent unnecessary harm and is therefore perceived to be ethical. Using both qualitative and experimental approaches, Levine (2022) established a number of systematic circumstances in which people judge deception as preventing unnecessary harm. For example, if a target is cognitively impaired (e.g., they suffer from dementia or are inebriated), if a target is under momentary duress (e.g., they are in a state of shock or grief), or if a target can no longer react to truthful information (e.g., there is no longer time to institute feedback), hurtful truths are perceived to cause unnecessary harm, and therefore, deception is perceived to be ethical. In general, as lies become more beneficial to the target of the lie, and less beneficial to the communicator, they are seen as more acceptable (Backbier, Hoogstraten, & Terwogt-Kouwenhoven, 1997). Lies can also be seen as acceptable when they are told out of necessity (e.g., to secure a job), or they are more subtle in nature (e.g., when a communicator says something true but misleading rather than something false; Rogers et al., 2017; Vrij, 2007).

Moral Judgments of Unconditional Honesty

Despite the belief that prosocial lies are often ethical, claims that "honesty is the best policy" abound (Huppert et al., 2023). Why would people endorse a policy that leads to suboptimal moral actions, namely the telling of harmful truths over prosocial lies? We propose that moral preferences for unconditional honesty stem from the unique desire to avoid *harmful* lies. We define unconditional honesty as abiding by a policy of telling the truth without considering the consequences of doing so. We compare unconditional honesty to conditional

honesty, which involves seeking out information about the consequences of honesty to inform one's decision to tell the truth.

Why unconditional honesty may be seen as ethical

Existing work on moral judgments of deception examines situations in which the consequences of deception are *known.* When people are certain that honesty will cause (unnecessary) harm, they believe that deception is ethical (Levine, 2022). However, we argue that when considering decisions under uncertainty, people believe that it is more important to avoid the most unethical actions than to pursue the most ethical ones. This proposition is consistent with existing research on risky decision making. Broadly, people are more attentive to avoiding negative outcomes than achieving positive outcomes when choosing gambles or other risky prospects (Kahneman & Tversky, 1979). This is true for risky moral decisions as well. For example, when making donation decisions, people are more likely to choose options that are guaranteed to help others rather than risky options, even when a risky option has higher expected (prosocial) value (Zlatev et al., 2020). Zlatev and coauthors (2020) theorize that "worst outcome avoidance" underlies this preference; people avoid prosocial risk in order to maintain their moral self-regard. If people act in a way that ensures they do *some* good, even if it is not the most possible good, they are able to legitimize their own moral standing. If people take prosocial risks that could lead to the worst outcome (e.g., not helping someone else), they risk losing that standing.

We theorize that a similar logic explains how people judge moral rules more broadly. Specifically, we theorize that people's judgments of moral rules are influenced by the degree to which following moral rules prevents the "worst outcome." In the context of honesty and deception, harmful lies are viewed as the "worst outcome." A careful examination of past work

is consistent with this view; although prosocial lies are seen as more ethical than harmful truths, harmful lies are seen as much more unethical than any other action. For example, Levine & Schweitzer (2014, Study 2) examine moral judgments of prosocial truths, selfish truths, prosocial lies, and selfish lies. They manipulated these acts using a deception game in which a communicator could either lie or tell the truth to earn $2 for themselves and $0 for a partner (harmful) or $1.75 for themselves and $1 for a partner (prosocial). Here, prosocial lies were seen as more moral than harmful truths. However, only harmful lies were seen as objectively unethical (significantly lower than 4, the midpoint of the 7-point rating scale of moral judgment; $M_{prosocial\ lie}$ = 4.80; $M_{harmful\ truth}$ = 4.31, $M_{prosocial\ truth}$ = 5.59; $M_{harmful\ lie}$ = 3.16). Similarly, Levine (2021; Study 1 – observer perspective) examined the percentage of people who believed lying (versus truth-telling) was ethical when the truth either would or would not lead to immediate harm and long-term instrumental value. When the truth would cause immediate harm and did not have instrumental value (and therefore prosocial lying would have no costs to a target in the long-run), a modest majority of participants (66.7%) believed that lying was more moral than truth-telling. In contrast, a large majority of participants believed that truth-telling was more ethical in all other situations (89.1% when the truth was high immediate harm, high instrumental value; 81.0% when the truth was low immediate harm, low instrumental value; 94.0% when the truth was low immediate harm, high instrumental value). This pattern of results provides further evidence that there is the greatest consensus that truth-telling is ethical and lying is unethical when lying would harm a target overall. Recent work on censorship also suggests that people view harmful lies as particularly dangerous (Kubin, con Sikorski, & Gray, 2022).

The hypothesis that people value unconditional honesty because it prevents the most unethical actions is also consistent with the broader literature on error management (Haselton &

Buss, 2000). Most decisions carry some risk of error, but these errors can vary from mild to extreme. Error management theory suggests that evolution favors decision rules that are predictably biased towards committing the less costly errors (Haselton & Buss, 2000). Applied to the present research, we argue that people endorse a policy of unconditional honesty because there is a cost asymmetry between harmful truths and harmful lies such that people believe harmful lies are more costly (unethical) than harmful truths. Unconditional honesty does allow for harmful truths, and therefore, does not always result in the most ethical choice. However, unconditional honesty prevents the *most* costly error: harmful lying. Unconditional honesty may not optimize every ethical decision, but it will minimize unethicality, overall.

The social rewards of unconditional honesty

If people believe that unconditional honesty is an effective strategy for avoiding the most unethical outcomes, they are likely to reward communicators who are unconditionally honest. People tend to make generalizations about others based on a single act (Gilbert & Malone, 1995). Therefore, people are likely to assume that someone who is unconditionally honest in a single situation is honest *in general*. Despite missing opportunities for prosocial lies, a person who is typically honest can be expected to avoid harmful lies across situations.

Conversely, someone who seeks out information about the consequences of truth-telling before deciding whether to tell the truth signals an openness to lying in general. Even if a communicator seeks out information with the goal of telling a prosocial lie, the communicator may nonetheless end up telling a harmful lie, due to the subjectivity of the consequences of lying (Vrij, 2007). For example, a communicator may try to figure out if a target could learn from hurtful, truthful information, intending to lie if the target could not use the information (Levine, 2022). However, this prosocially-motivated communicator may be biased in their assessment

about whether the target could learn, leading them to lie to the target, when the truth would have actually been more beneficial. Indeed, paternalistic beliefs and self-other perspective gaps lead people to underestimate the benefits and overestimate the harm associated with truth-telling in feedback settings and beyond (Abel et al., 2022; Abi-Esber, Abel, Schroeder, & Gino, 2022; Levine & Cohen, 2018; Lupoli, Levine, & Greenberg, 2018). Communicators could also be conditionally honest for politeness (Brown & Levinson, 1987), conflict-avoidance (De Paulo et al., 1996) or impression-management (Goffman, 1955) reasons, and end up telling lies that are harmful to targets, despite the communicator not having harmful intentions.

Seeking out information about the consequences of honesty might also sway a person to lie for reasons beyond prosociality. Some communicators may seek out information about the consequences of honesty because they explicitly want to tell harmful, selfish lies. In contrast, other communicators may seek out information about the consequences of honesty with the intention of promoting prosociality but then find it hard to resist the temptation to pursue self-interest after realizing that harmful lying would be personally beneficial. Observers may recognize that looking carries risks for communicators who have a weakness of will, whereas unconditional honesty is a safeguard against lies of all kinds. Overall, we propose that engaging in unconditional honesty is socially rewarded because it signals that a communicator generally avoids harmful lies, which is uniquely associated with positive moral character.

This proposition dovetails with recent work, documenting the reputational benefits of taking absolute stances on honesty. In particular, Huppert and colleagues (2023) find that observers are more willing to trust and support (i.e., vote for) communicators who take absolute honesty stances (e.g., “It is never okay to lie) relative to communicators who take nuanced stances (e.g., “It is sometimes okay to lie), even when both communicators lie. Communicators –

including both lay people and political leaders – seem to anticipate these effects, and therefore, are more likely to take absolute stances on honesty in public than in private. Policies of absolute honesty are viewed positively because they are seen as genuinely informative about one's overall commitment to honesty and predictive of future choices. Similarly, people who engage in unconditional honesty are likely to seem genuinely committed to future honesty, leading to less harmful lying, which we argue is more important to observers than missed opportunities for prosocial lies. Furthermore, unconditional honesty helps promote a norm of truth-telling (Huppert et al., 2023), which might encourage moral rules and curb harmful lying in society broadly. Therefore, by signaling information about a communicator's future truth-telling and promoting truth-telling in others, unconditional honesty should help minimize harmful lying overall.

Overview of Research

Empirical Approach

We test our hypothesis across five main experiments (as well as 10 supplemental studies), using both economic games and face-valid vignettes. We build on the "cooperation-without-looking" (Hoffman et al., 2015; Jordan et al., 2016) paradigm to compare moral judgments of two different strategies that people could pursue when faced with dilemmas involving honesty: 1) "unconditional honesty," which binds communicators to truth-telling, regardless of whether the truth ultimately helps or harms a target, and 2) "looking" (i.e., conditional honesty), which allows people to condition their honesty on its consequences. We use the terms "looking" and "conditional honesty" interchangeably in the remainder of the manuscript.

In our studies, we focus on situations in which honesty has consequences for *others*, rather than the communicator. We presumed that seeking out information about the personal

consequences of honesty would be penalized, given the reputational costs of selfishness (Berman & Silver, 2022) and the benefits of signaling commitment and decision-certainty when facing conflicts between morality and self-interest (Critcher et al., 2013; Jordan et al., 2016). Indeed, we find evidence that seeking out information about the personal costs of honesty is penalized (see Studies S2 and S3 in SOM 2.2-2.3). However, seeking out information about the *social costs* of honesty captures the tension of theoretical interest – it allows a communicator to tell prosocial lies, but also opens up the possibility of harmful lies.

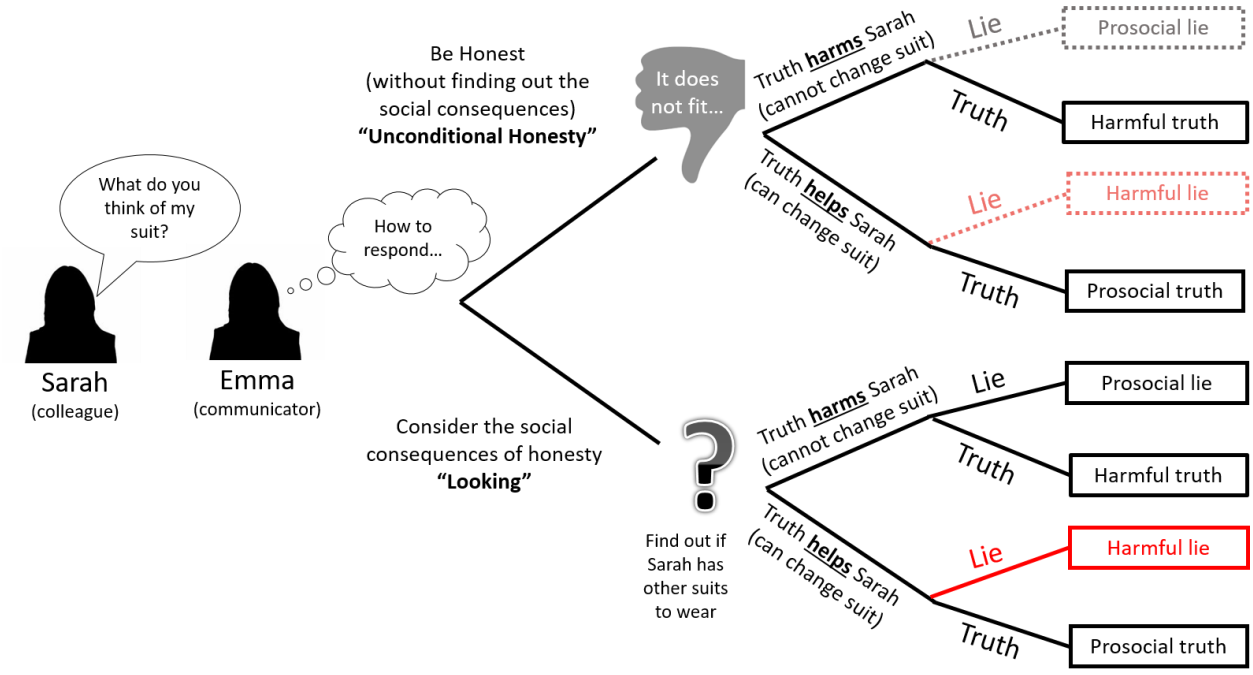
We illustrate unconditional and conditional honesty within the context of providing feedback in Figure 1. Imagine that a person named Sarah has an important presentation at work. She asks her colleague, Emma, for her opinion on the suit she intends to wear. Emma believes the suit is ill-fitting and unprofessional. What do people believe Emma should do? Prior research suggests that knowing whether Sarah has other suits she can change into (i.e., whether Sarah can effectively use the feedback; Levine, 2022) is material to judgments of whether it is ethical for Emma to tell Sarah her true opinion. If Sarah cannot change her suit, knowing that the suit is ill-fitting causes unnecessary harm, and lying is seen as both prosocial and ethical. However, if Sarah *can* change her suit, knowing the suit is ill-fitting allows Sarah to change into something more appropriate, and lying to Sarah would be harmful, as well as unethical.

Rational decision theory posits that people should seek out information that is material to a decision and costless to receive. Therefore, if people believe that the moral response depends on whether Sarah can change her suit, they should also believe that Emma is morally obligated to find out this information (i.e., to "look" for this information). However, across five main experiments, we find that people believe unconditional honesty is a more ethical decision strategy than looking and reward communicators who engage in unconditional honesty.

Although looking allows communicators to condition their decision on the social consequences of honesty, unconditional honesty prevents the worst outcome – harmful lies – from occurring.

Figure 1

Unconditional honesty versus Looking Decision Tree



Notes. The decision tree specifies the outcomes associated with two of the communicator's decision strategies (Unconditional Honesty vs. Looking). Unconditional Honesty binds the communicator to truth-telling, which could lead to a harmful truth or a prosocial truth. Looking gives the communicator the ability to condition their choice on the social consequences of honesty, but also allows for harmful lies (the most unethical action, represented by red outline).

In Study 1, we examine moral judgments of unconditional honesty using the same thought experiment described in Figure 1 (providing feedback on an ill-fitting suit). Then, in Studies 2-5, we examine the social and reputational consequences of unconditional honesty. We examine how engaging in unconditional honesty influences judgments of communicators' moral character, and reliance on communicators' advice. We also demonstrate that concerns about a

communicator's propensity to tell harmful lies underlies moral judgments of unconditional honesty.

Transparency and Openness

Across our studies, our stopping rules for data collection were decided in advance, and we report all measures and conditions we collected. In Study 1, 2, 3, 4, and 5 respectively, we had post-hoc powers of .82, .56, .99, .72, and .99 to detect our focal effects on morality. In Study 1, we used the observed *z* statistics for the focal effect of the proportion selecting Unconditional honesty as the most ethical decision strategy when the consequences of honesty were known versus unknown to compute power using a post-hoc power analysis for independent-samples z-tests comparing two proportions, with $\alpha = .05$ in G*power. In Studies 2- 5, we used the observed Cohen's *d*s for the focal effect of Unconditional Honesty versus Looking on perceived morality and used the sample sizes of the relevant conditions for each study to compute power using a post-hoc power analysis for independent-samples t-tests with $\alpha = .05$ in G*power, version 3.1 (Faul et al., 2007). For Studies 3 and 4, we also based our sample sizes on a priori power analyses informed by pilot tests, which we report in the methods section of each study.

Our reported samples consist of all participants who correctly answered the study comprehension checks and completed the study in its entirety. Participants who did not correctly answer comprehension checks within two attempts were directed to the end of the survey without completing our dependent variables, and therefore, are not included in any analyses. Studies 2, 3, 4, and 5 were preregistered on aspredicted.org. Across our studies, we report the results of our main, preregistered analyses, which did not correct for multiple comparisons. All data, syntax, and materials are available at: https://tinyurl.com/HWOL-OSF. The Institutional Review Board at the University of Chicago approved all studies. At the end of each study, participants reported

their demographic information from a provided set of options. In Studies 1-3, participants selected which option best describes themselves (options: male, female, prefer to self-describe, prefer to not answer). In Studies 4 and 5, participants selected their gender (options: man, woman, prefer to self-describe, prefer to not answer).

Study 1: Moral Judgments of Unconditional Honesty

In Study 1, we examine the moral preference for unconditional honesty using a face-valid vignette. In doing so, we show that the preference for unconditional honesty cannot be explained by the belief that information about the social consequences of honesty is irrelevant to moral judgments. Importantly, we also replicate existing work by showing that although people believe that prosocial lies are more ethical than harmful truths, they have the strongest belief that harmful lies are unethical.

Method

Participants. We recruited 199 participants (M_{age} = 35.87, SD_{age} = 11.11; 120 males, 79 females) from Amazon Mechanical Turk (MTurk).

Procedure and materials. Participants read a scenario similar to the thought experiment corresponding to Figure 1; they read about a situation in which an employee who was about to give an important presentation asked a colleague how s/he looked in an ill-fitting suit (adapted from Levine, 2022). Participants were randomly assigned to one of two conditions in a between-subjects design; Participant either knew or did not know the social consequences of being honest (Consequences: Known vs. Unknown).

Participants in the Known condition were presented with two versions of the same scenario in a randomized order and made a judgment within each version of the scenario. Specifically, participants indicated whether lying or truth-telling would be the most ethical

choice if the colleague were certain that honesty harmed the employee (i.e., if the employee could not change their suit) *and* if the colleague were certain that honesty helped the employee (i.e., if the employee could change their suit). For each judgment, participants indicated whether the most ethical decision for the colleague to make was: a) "Tell the employee the truth—say that he thinks the suit is inappropriate" or b) "Lie to the employee—say he thinks the suit is fine." The exact text from each condition is reported in SOM 1.1.1.

Participants in the Unknown condition made a single judgment, without knowing the social consequences of honesty. Specifically, they did not know whether the employee could change their suit. As a result, participants did not know whether providing honest critical feedback would ultimately help or harm the employee. Participants indicated whether the most ethical decision for the colleague to make was: a) "Tell the employee the truth — say that he thinks the suit is inappropriate" (this represents Unconditional Honesty), b) "Lie to the employee — say that he thinks the suit is fine" (this represents Unconditional Lying), or c) "find out if the employee owns another suit before answering" (this represents Looking).

Analytical Approach

We used the Known condition to replicate past work, and ensure we were studying a situation in which people believe that prosocial lying is ethical. In line with existing work (Levine, 202), we found that a modest majority (55%) of participants believed that lying was the most ethical choice when honesty was harmful and lying was prosocial. In contrast, when honesty was prosocial and lying was harmful, the majority of participants (78%) believed honesty was the most ethical decision.[1] Consistent with our account, there seemed to be greater

[1] Notably, a fair amount of people still believed telling the harmful lie was ethical. In this scenario, the harmful lie is polite in the moment (i.e., "the suit looks fine"), but harmful in the long-run because it prevents the target from changing into a more appropriate suit. Although the

consensus that the harmful lie rather than the harmful truth was unethical, lending support to the notion that harmful lies are generally viewed as the most unethical actions.

In the Known condition, we also calculated the percentage of people who conditioned their moral judgments on the social consequences of honesty (i.e., by making different moral judgments when honesty helped vs. harmed the employee). Changing one's judgment based on the social consequences of honesty suggests that the social consequences are material to moral judgments, implying that Looking is the most ethical strategy. According to rational decision theory, if information is free and relevant to a decision, people *should* seek it out. Therefore, the percentage of people who condition their moral judgments on the social consequences of honesty in the Known condition should be roughly equal to the percentage of people who believe it is moral to find out information about the social consequences of honesty in the Unknown condition (i.e., those who choose Looking).

Our main planned analysis was to compare these percentages; that is, we intended to compare the distribution of choices (Unconditional Honesty, Unconditional Lying, Looking) expressed in the Unknown condition to the distribution of choices implied by the set of decisions in the Known condition (endorsing honesty in both contexts implies Unconditional Honesty, endorsing lying in both contexts implies Unconditional Lying, and making different decisions across contexts implies Looking).

Results

The distribution of choices in the Unknown condition was significantly different than the distribution of implied choices in the Known condition, $X^2 = 6.70$, $p = .035$. In the Unknown

---

majority of people tend to believe truth-telling is ethical in these situations, it is rarely a unanimous opinion (Levine, 2022).

Condition, the majority (58%) of participants chose Unconditional Honesty as the most ethical decision strategy, compared to just 40% of participants in the Known condition who endorsed honesty across both situations they faced (implying a preference for Unconditional Honesty). In the Unknown condition, only 29% of people chose Looking (13% of people chose Unconditional Lying), compared to the 44% of participants who conditioned their moral judgments on the consequences of honesty in the Known condition, implying a preference for Looking (16% of participants indicated that lying was the most ethical choice in both situations, implying a preference for Unconditional Lying).

Discussion

Study 1 reveals that most people believe that unconditional honesty is the most ethical decision strategy when the communicator does not know whether honesty will help or harm a target. People believe it is ethical to avoid finding out whether honesty will cause harm, and instead, to simply tell the truth. However, using the same paradigm, we also show that when people do know whether honesty harms or helps others, people base their ethical beliefs on this information. These results suggest that the moral preference for unconditional honesty does not simply stem from a belief that information about the social consequences of honesty is immaterial to moral judgments.

These results are consistent, however, with our hypothesis that moral judgments are negatively associated with the belief that a communicator will tell harmful lies. Indeed, a post-test of Study 1 revealed that negative moral judgments of conditional honesty (i.e., Looking) in this paradigm were associated with the belief that a communicator who was conditionally honest would tell lies that harm others (b = -0.22, SE = 0.07, $p < .001$; see SOM 1.1.2 for details).

We ran a preregistered conceptual replication of Study 1, in which we use a within-subjects, rather than between-subjects, design, which we report in the supplemental online materials (see SOM 2.1). We replicate our results within-subjects, suggesting that the preference for unconditional honesty cannot be explained by a failure to think through the relevant decision tree (Shafir & Tversky, 1992). People believe that unconditional honesty is ethical when the social consequences of honesty are unknown, even after they make judgments about what is most ethical when the consequences are known. In other words, people believe that communicators should *not* seek out information that they believe is necessary for making morally optimal decisions.

Study 2: The Reputational Consequences of Unconditional Honesty

In the remainder of our studies, we examine the social and reputational consequences of unconditional honesty. In Study 2, we examine moral judgments of, and the willingness to rely on advice from, communicators who are either honest without looking at the social consequences of honesty (Unconditional Honesty), or who look at social consequences of honesty before deciding whether or not to tell the truth (Looking) when playing an economic game.

Method

Participants. As preregistered, we aimed to recruit as many participants as possible in one day from a virtual research laboratory at the University of Chicago Booth School of Business. We ended up with a final sample of 240 participants who passed initial comprehension checks and were eligible to complete the full survey (M_{age}= 27.39, SD_{age} = 9.57; 57 males, 178 females, 4 prefer to self-describe, 1 prefer to not answer).

Procedure and materials. We randomly assigned participants (between-subjects) to interact with a Communicator who had either engaged in Unconditional Honesty or Looking (Decision Strategy: Unconditional Honesty vs. Looking).

Participants were paired with a Communicator that previously participated in a study in which they played the Coin Flip Game (adapted from Levine & Munguia Gomez, 2021). Participants learned the rules of the Coin Flip Game that the Communicator played.

In the Coin Flip game, Communicators flipped a digital coin and then had to report the outcome of the coin flip to the experimenter. Communicators learned that their report would influence the payment of a partner, the Target, but the exact monetary values associated with their report were uncertain. Specifically, Communicators learned:

- If they reported that the coin landed on HEADS, the Target would receive $A
- If they reported that the coin landed on TAILS, the Target will receive $B

Although the exact amounts were uncertain, Communicators did learn information about the possible values of A and B. Communicators knew that there was a 50% chance that $A was $1 and $B was -$1, and a 50% chance that $A was -$1 and $B was $1. As a result, the probabilities that telling the truth would help the Target, telling the truth would harm the Target, telling a lie would help the Target, and telling a lie would harm the Target were all equal.[2]

[2] This design feature helps to address the possibility that people think lying is more likely to cause harm than truth-telling is. If people do not believe that truth-telling leads to harm particularly often, but that lying does, a policy of Unconditional Honesty minimizes overall harm. This alternative account is consistent with the "social heuristics hypothesis" (Rand, et al., 2014), which suggests that because cooperation is typically advantageous, we have developed intuitive heuristics that favor it that are deployed even in settings where cooperation is no longer advantageous (i.e. one-shot anonymous interactions in the laboratory). Similarly, people may believe that honesty is typically associated with welfare-maximization, leading people to favor honesty as a general heuristic. Though this mechanism may also be at play in certain contexts,

Next, the Communicator flipped the coin. The Communicator then made one of three possible choices: They either selected "The coin landed on HEADS," "The coin landed on TAILS," or "I'd like to find out the values of $A and $B before making my decision." Roughly half of the participants in our study learned about a Communicator who honestly reported the outcome of the coin flip without looking at the consequences associated with the decision to tell the truth (Unconditional Honesty condition). The other half of participants learned about a Communicator who chose to look at the consequences of telling the truth before reporting the outcome (Looking condition). Although unconditional lying was also a choice, no Communicators were described as engaging in Unconditional Lying, as this was not our focal interest. Figure 2 depicts the design of the study.

After learning about the Communicator's behavior in the Coin Flip Game, participants played the Advice Game with the Communicator. The Advice Game was adapted from the Weight of Advice Task used in Gino & Schweitzer (2008). In this task, participants made an initial estimate of how much money was in a jar of coins. Then, participants received advice from the Communicator about how much money was in the jar. Participants knew that the Communicator knew the true amount of money in the jar, but they did not know whether the Communicator was incentivized to give truthful or untruthful advice to the participant. After seeing the Communicator's advice, participants had the option to revise their guess before reporting the final estimate of how much money they thought was in the jar of coins. If a participant's final estimate was within $1 of the true amount of money in the jar of coin, the participant received a bonus.

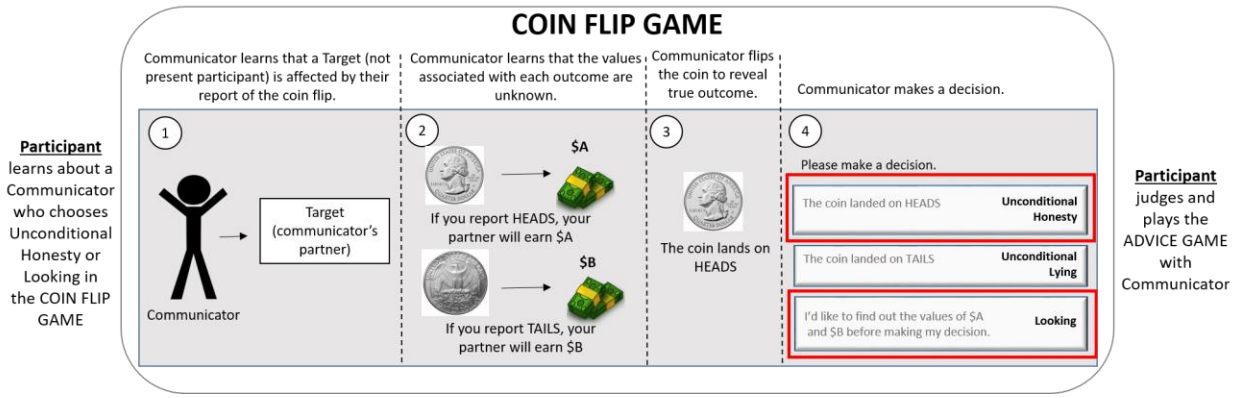
---

we control for these inferences in our studies by clarifying that truth-telling and lying were equally likely to lead to prosocial versus harmful outcomes.

We calculated weight of advice (WOA) by taking the absolute value of the difference between the participant's final and initial estimates and dividing it by the absolute value of the difference between the amount of the Communicator's advice and the participant's initial estimate. WOA captures the degree to which the participant's final estimate is anchored on their own initial estimate versus the advice they received from the Communicator; higher numbers reflect greater reliance on the Communicator's advice. In other words, WOA is a continuous measure of the degree to which a person trusts the Communicator's advice. After playing the Advice Game, participants judged the Communicator's morality using a three-item composite of moral, good, and ethical (1= Not at all, 7= Extremely; α = .944; adapted from Effron & Monin, 2010; Levine & Schweitzer, 2014, Uhlmann, Zhu & Tannenbaum, 2013).[3]

Figure 2

Design of the Coin Flip Game in Study 2



Notes. The figure depicts the experimental design in which participants judged and interacted with communicators that either used the decision strategy of Unconditional Honesty or sought out more information (Looking) within the Coin Flip Game. Participants did not learn what happened next if Communicators chose Looking.

[3] We also included a measure of perceived trustworthiness, which showed the same directional pattern as WOA but was not significant. The details of all ancillary measures in our studies are reported in the SOM.

Results

Consistent with our preregistration, we conducted independent samples t-tests on morality and WOA, using Decision Strategy as the independent variable. We find that participants relied on the advice of the Communicator who was unconditionally honest (M = 0.68, SD = 0.66) more than the Communicator who looked at the consequences of honesty (M = 0.52, SD = 0.42; $t(238) = 2.37$, $p = .019$). Participants also judged Communicators who were unconditionally honest as significantly more moral ($M = 4.37$, $SD = 1.14$) than those who looked at the consequences of honesty ($M = 4.06$, $SD = 1.15$; $t(238) = 2.12$, $p = .035$). Effect sizes (Cohen's d) associated with key contrasts in Study 2 and all other studies with a similar design are reported in Appendix B (see Table A1).

Discussion

Study 2 provides evidence that people judge communicators who engage in unconditional honesty as more moral than communicators who engage in looking (i.e., conditional honesty) and are more likely to rely on the advice of communicators who are unconditionally honest. Even though our paradigm clarified that honesty was just as likely to harm as to help a communication partner, and that this uncertainty could have been resolved by choosing to find out the values of $A and $B, people believed that communicators who did *not* seek out this information were more moral.

Study 3: Harmful Lies and the Moral Preference for Unconditional Honesty

In Study 3, we extend our investigation in three ways. First, we measure our proposed mechanism: beliefs about the degree to which communicators engage in harmful lies. Second, we rule out default effects as an explanation of our results. In Study 2, communicators who chose Looking had to actively choose to find out more information. It is possible that in this design,

unconditional honesty was perceived to be a default strategy, and therefore seen as a more normative option (Krijnen, Tannenbaum, & Fox, 2017). In Study 3, we change the perceived default in the Coin Flip Game to rule out this explanation. Specifically, in Study 3, communicators automatically learned the consequences of their honesty before deciding whether to tell the truth or lie *unless* they chose to avoid learning the consequences. Third, we compare observer judgments of communicators to the actual behavior and self-reported morality of communicators to determine the accuracy of observer judgments.

Method

Participants. As preregistered, we recruited 450 participants using the Academic Prolific platform. Of those participants, 444 participants provided demographic information (M_{age} = 36.25, SD_{age} = 12.23; 220 men, 217 women, 7 prefer to self-describe, 6 did not answer). Our sample size was based off an a priori power analysis informed by a small pilot (N = 76) using the same design as Study 3. In this pilot, the effect size (comparing Unconditional Honesty to Looking) for moral character was approximately d = .242. This revealed we needed to recruit a sample of N = 434 participants to achieve 80% power. Thus, we preregistered and recruited a sample of 450 participants.

Procedure and materials. Participants were randomly assigned to one of two experimental conditions (Decision Strategy: Unconditional Honesty vs. Looking) in a between-subjects design. Participants made judgments of a Communicator who previously played the Coin Flip Game, similar to the one we used in Study 2. However, we made two key changes to our design in Study 3.

First, we changed the perceived default in the game. In Study 3, we constructed the Coin Flip Game so that the default decision strategy was Looking. Specifically, Communicators

flipped the coin, and learned the outcome of the coin flip. Then, Communicators learned that the values of $A and $B would be revealed to them before they were asked to report the outcome of the coin flip *unless* they chose the option, "I do not want to find out the values of $A and $B before making my decision." Participants observed a Communicator who either chose not to find out the values of $A and $B and then honestly reported the outcome of the coin flip (Unconditional Honesty) or a Communicator who proceeded with the Game and learned the values of $A and $B (Looking). In this design, the decision to engage in Unconditional Honesty by avoiding additional information was clearly an active, intentional choice.

Second, we collected additional measures to examine the process by which Unconditional Honesty leads to positive judgments of moral character. Specifically, we measured the extent to which participants believed the Communicator: (i) would tell a harmful lie in the Coin Flip Game, (ii) would tell a prosocial lie in the Coin Flip Game, (iii) has the tendency to tell harmful lies in general, and (iv) has the tendency to tell prosocial lies in general.

Third, we compared the judgments and behaviors of actual Communicators that we recruited in a separate study to observer evaluations of Communicators in Study 3 to examine whether observers' inferences about Communicators reflect reality. In doing so, this study sheds light on whether positive judgments of unconditionally honest communicators are *accurate*.

Dependent Variables.

Morality. Our primary dependent variable was judgments of the Communicator's morality using the same three item composite used in Study 2 ($\alpha = .923$).

Moral Identity. We measured judgments of moral identity. We asked actual Communicators to self-report their moral identity as well, so the goal of including this additional

observer judgment of morality was to compare observers' judgments of Communicator's moral identity to Communicator's self-ratings.

Following prior measurements of moral identity (Reed & Aquino, 2003), participants read: "Listed below are some characteristics that might describe a person: Caring, Compassionate, Fair, Friendly, Generous, Helpful, Hardworking, Honest, and Kind. The person with these characteristics could be you or it could be someone else. For a moment, visualize in your mind the kind of person who has these characteristics. Imagine how that person would think, feel, and act. When you have a clear image of what this person would be like, answer the following questions." Then, participants reported judgments of the Communicator's moral identity using four-items on a 1 (not at all) to 7 (very much so) scale: "It would make [the Communicator] feel good to be a person who has these characteristics.", "Being someone who has these characteristics is an important part of who [the Communicator] is.", "[The Communicator] would be ashamed to be a person who had these characteristics (*reverse coded*).", and "Having these characteristics is not really important to [the Communicator] (*reverse coded*)."

Probability of telling a Harmful or Prosocial Lie in the Coin Flip Game. Participants reported their belief that the Communicator would tell harmful and prosocial lies within the Coin Flip Game, and in general. Participants reported their beliefs about harmful lies in the Coin Flip Game using one item scale (0% - 100%): "How likely is it that the DECIDER you learned about would lie and cause the RECEIVER to get $B, If $A is $1, $B is -$1?" Participants also reported their belief that the Communicator would tell a prosocial lie in the Coin Flip Game using one item scale (0% - 100%): "How likely is it that the DECIDER you learned about would lie and cause the RECEIVER to get $B, If $A is -$1, $B is $1?."

Tendency to tell Harmful and Prosocial Lies. Participants also reported their belief that the Communicator would tell harmful lies and prosocial lies in general using two items on a 1 (not at all) to 7 (a great extent) scale: "Outside of the context of the Coin Flip Game, to what extent do you think [the Communicator] would tell lies that harm others?" and "Outside of the context of the Coin Flip Game, to what extent do you think [the Communicator] would tell lies that help others?".

Results

Morality. Communicators who engaged in Unconditional Honesty (M = 5.48, SD = 1.22) were viewed as significantly more moral than those who engaged in Looking (M = 4.68, SD = 1.03; $t(448) = 7.49$, $p < .001$).

Moral Identity. Communicators who engaged in Unconditional Honesty (M = 5.71, SD = 0.94) were also viewed as being significantly higher in moral identity than those who engaged in Looking (M = 5.35, SD = 0.95; $t(447) = 4.03$, $p < .001$). These results are depicted in Figure 3, Panel A.

Probability of telling a Harmful or Prosocial Lie in the Coin Flip Game. Communicators who engaged in Unconditional Honesty (M = 15.37, SD = 26.52) were judged as significantly less likely to tell a harmful lie in the Coin Flip Game than those who engaged in Looking (M = 24.74, SD = 25.15; $t(439) = 3.80$, $p < .001$). Communicators who engaged in Unconditional Honesty were also judged as significantly less likely to tell a prosocial lie in the Coin Flip Game (M = 18.96, SD = 29.14 vs. M = 48.47, SD = 28.25; $t(434) = 10.74$, $p < .001$). These results are depicted in Figure 3, Panel B.

Tendency to tell Harmful and Prosocial Lies in General. Communicators who engaged in Unconditional Honesty (M =1.90, SD = 1.00) were seen as less likely to tell harmful

lies (in general) than Communicators who engaged in Looking (M = 2.59, SD = 1.21; $t(442)$ = 6.54, $p < .001$). Communicators who engaged in Unconditional Honesty were also seen as less likely to tell prosocial lies in general (M = 3.17, SD = 1.56 vs. M = 4.16, SD = 1.40; $t(442)$ = 7.03, $p < .001$).

Mediation Analyses. We conducted a mediation model with 10,000 samples using Decision Strategy as the independent variable (1 = Unconditional Honesty, 0 = Looking), judgments of the Communicator's morality as the dependent variable, and the probability of telling a harmful lie and prosocial lie in the Coin Flip Game, as well as the tendency to tell harmful and prosocial lies in general as simultaneous mediators (PROCESS Macro for SPSS, Model 4; Hayes, 2013).

We found evidence of significant mediation through the tendency to tell harmful lies in general (indirect effect = 0.23, SE = 0.05, 95% CI [0.14, 0.32]). Communicators who engaged in Unconditional Honesty were judged as having a lower tendency to tell harmful lies (b = -0.70, $p < .001$) than those who engaged in Looking, and the perceived tendency to tell harmful lies was negatively related to judgments of morality (b = -0.32, $p < .001$). We also found significant mediation through the tendency to tell prosocial lies in general, although the effect was notably smaller (indirect effect = 0.09, SE = 0.04, 95% CI [0.08, 0.17]). Communicators who engaged in Unconditional Honesty were judged as having a lower tendency to tell prosocial lies (b = -1.00, $p < .001$) than those who engaged in Looking, and the perceived tendency to tell prosocial lies was negatively related to judgments of morality (b = -0.09, p = .021), albeit to a lesser extent than the perceived tendency to tell harmful lies.

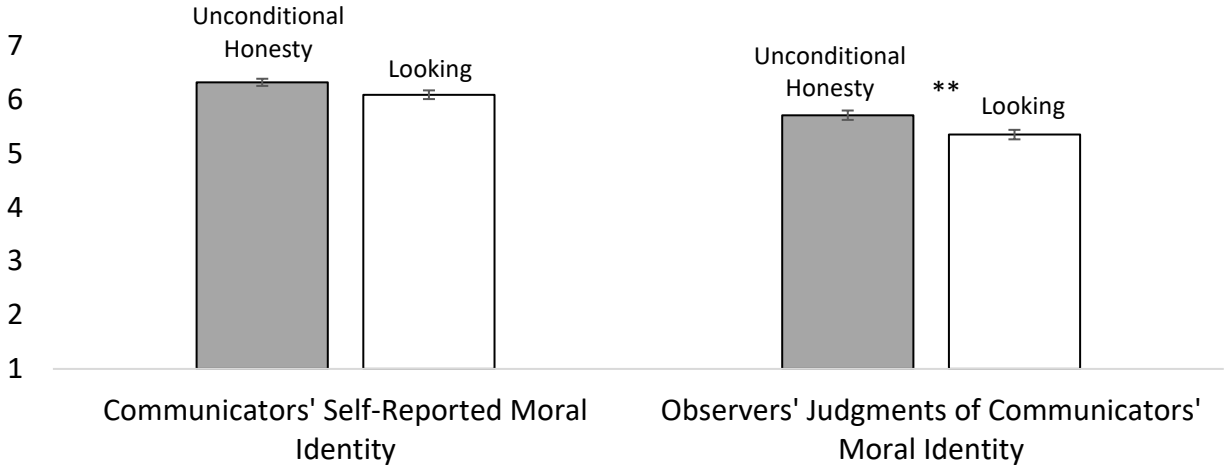
We did not detect significant mediation through judgments of the probability that the Communicator would tell a harmful lie in the Coin Flip Game (indirect effect = 0.01, SE = 0.02,

95% CI [-0.04, 0.05]), nor the probability that the Communicator would tell a prosocial lie in the Coin Flip Game (indirect effect = -0.08, SE = 0.06, 95% CI [-0.19, 0.03]). These results suggest that Unconditional Honesty is beneficial for one's reputation because it signals that one is unlikely to engage in harmful lies in general, above and beyond judgments about harmful lies within a particular context. We find similar mediation results when we use judgments of moral identity as our dependent variable.

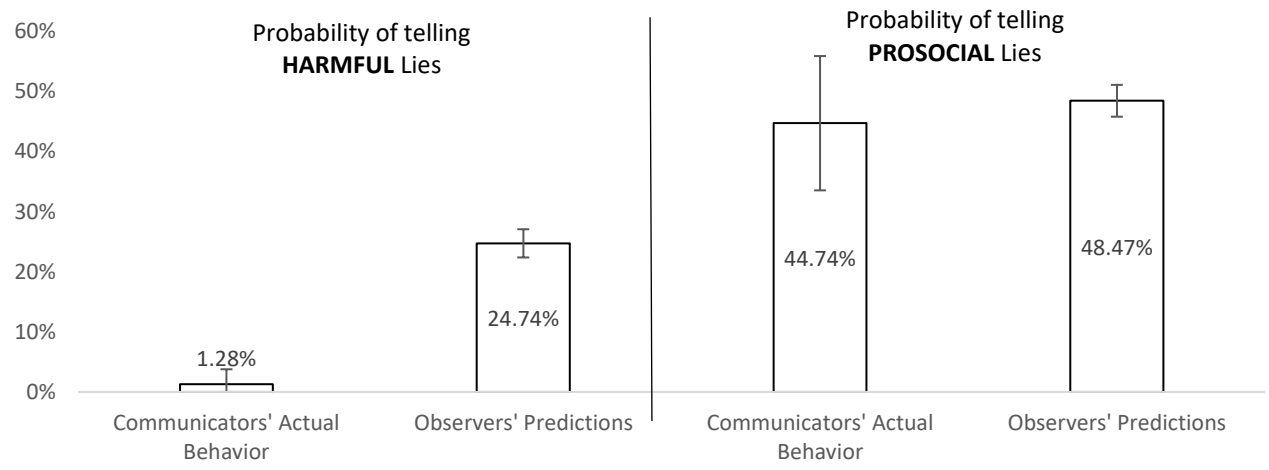
Figure 3

Moral Identity Judgments and Beliefs about Harmful Lies in Study 2

Panel A. Communicators' self-reports of Moral Identity vs. Observers' judgments of Communicators' Moral Identity as a function of Communicators' decision to engage in Unconditional Honesty vs. Looking



Panel B. Observers' predictions and Communicators' actual likelihood of engaging in harmful and prosocial lies after Looking



Notes. In Study 3, Observers believed that Communicators who engaged in Unconditional Honesty were higher in moral identity than Communicators who engaged in Looking, though Communicators' self-reports did not align with these judgments (Panel A). Observers believed that Communicators who engaged in Looking were more likely to tell harmful lies than Communicators actually were. Observers were fairly well-calibrated to the likelihood of telling prosocial lies. Error bars represent 95% confidence intervals around the mean. **signifies $p < .001$.

Are these judgments accurate? To explore whether judgments of unconditionally honest Communicators correspond to their self-perceived moral identity, we ran a separate sample of Communicators. In this study, Communicators were asked to complete the Coin Flip Game and self-report judgments of their own moral identity. We examined the likelihood that Communicators who engaged in Looking actually told harmful lies and compared the self-rated moral identity of Communicators who were Unconditionally Honest to Communicators who engaged in Looking.

Of the 189 Communicators (M_{age} = 35.48, SD_{age}= 12.14; 92 men, 93 women, 3 prefer to self-describe, 1 prefer to not answer), 35 (18.5%) chose to report the outcome of the coin flip without Looking (i.e., selected "I do not want to find out the values of $A and $B before making my decision"). The remaining 154 (81.5%) engaged in Looking and chose to view the values of

$A and $B before reporting the outcome of the coin flip. The fact that a large majority of Communicators actually chose to Look in this paradigm suggests that Unconditional Honesty is not rewarded simply because it is the choice most participants would make for themselves.

We randomly assigned the Communicators who engaged in Looking to learn that the truth would harm the RECEIVER (i.e., $A is worth -$1, and $B is worth $1) or to learn that the truth would help the RECEIVER (i.e., $A is worth $1, and $B is worth -$1). Of the 76 Communicators who learned the truth would harm the RECEIVER, 42 (55.3%) told the harmful truth and 34 (44.7%) told the prosocial lie. Of the 78 Communicators who learned the truth would help the RECEIVER, 77 (98.7%) told the prosocial truth and 1 (1.3%) told the harmful lie. These results suggest that observers may overestimate the likelihood that Communicators who Look for more information will tell harmful lies; participants in the main study estimated the probability of telling a harmful lie after Looking was 24.74%, but in reality it was only 1.3% (see Figure 3, Panel B).

We also conducted an independent samples t-test on moral identity using Communicators' chosen Decision Strategy in the Coin Flip Game as an independent variable to compare whether Communicators who engage in Unconditional Honesty see themselves as more or less moral than Communicators who engage in Looking. We did not detect an effect of Decision Strategy on self-reported moral identity (p = .151; see Figure 3, Panel A). Communicators who engaged in Unconditional Honesty (M = 6.32, SD = 0.72) viewed themselves as possessing similar moral identity as Communicators who engaged in Looking (M = 6.09, SD = 0.87).

Discussion

Replicating the results of Study 2, communicators who engaged in unconditional honesty were judged as more moral than those who engaged in looking. Importantly, Study 3 reveals that unconditional honesty yields more positive moral judgments, even when unconditional honesty reflects an intentional choice to avoid information. Furthermore, Study 3 explores the mechanisms underlying these judgments. Although communicators who engaged in looking were judged as more likely to tell prosocial lies and harmful lies in the Coin Flip Game, and to tell prosocial lies in other contexts, our mediation results suggest that looking is penalized relative to unconditional honesty primarily because it signals that a communicator is more likely to tell harmful lies across contexts. This inference is uniquely damaging to judgments of moral character, consistent with our theoretical account.

Study 3 also provides initial evidence that judgments of communicators who look for more information may be miscalibrated. Participants in Study 3 overestimated the degree to which communicators who looked for information ultimately told harmful lies and underestimated their moral identity.

Study 4: The Moral Preference for Unconditional Honesty Across Relationships

In Study 4, we extend our investigation by examining how unconditional honesty influences a range of relationship preferences and interpersonal behaviors. We also operationalize unconditional honesty in a new way, using a survey exchange paradigm.

Method

Participants. As preregistered, we aimed to recruit 450 participants from the University of Chicago virtual laboratory. We ended up with a final sample of 448 participants (M_{age} = 29.99, SD_{age} = 11.20; 31.5% men, 65.8% women, 2.2% prefer to self-describe, 0.4% prefer to not answer). Our sample size was based off an a priori power analysis that was informed by a small

pilot (N = 138). In this pilot study, we compared Unconditional Honesty to Looking, and the effect size for moral character was d = .315. In order to achieve 80% power, it was revealed that we would need a sample of N = 320. In order to achieve sufficient power while being consistent with Study 3, we aimed to recruit 450 participants.

Procedure and materials. Participants were randomly assigned to one of two experimental conditions (Decision Strategy: Unconditional Honesty vs. Looking) in a between-subjects design. Across conditions, participants learned that they would be paired with a (fictitious) past participant (who we call the "Communicator"). Participants saw a screenshot of a question the Communicator had been presented with in a previous study, which featured a scenario about giving feedback. Below the question, participants saw the Communicator's response, which we altered based on the condition. Participants either read that the Communicator would adopt a policy of unconditional honesty or that they would look for information before giving feedback to a coworker (see Figure A1 in Appendix A for exact text). After reading the Communicator's response, participants judged the Communicator. At the end of the study, participants were debriefed.

Dependent Measures

Morality. We measured morality using the same three item composite used in Studies 2 and 3 (α = .906).

Tendency to tell harmful and prosocial lies. Participants reported their belief that the Communicator would tell both harmful lies and prosocial lies outside of the present context, using similar items as those used in Study 3.

Relationships. We asked participants to rate their agreement about whether they would want their partner as a friend, manager, coworker, and leader (1 = Strongly disagree, 7 = Strongly agree).

Behavioral Intentions. We also asked participants, "How frequently would you trust the advice of [your partner]?" and "How frequently would you ask [your partner] for advice if you needed it?" (1 = Never, 7 = Always).

Results

Morality. Communicators who were Unconditionally Honest (M = 6.00, SD = 0.86) were rated as more moral than Communicators who Looked (M = 4.89, SD = 1.27; $t(446) = 10.87$, $p < .001$).

Tendency to tell Harmful and Prosocial Lies in General. Communicators who engaged in Unconditional Honesty (M =1.79, SD = 1.07) were rated as less likely to tell harmful lies (in general) than Communicators who engaged in Looking (M = 2.58, SD = 1.51; $t(446) = -6.40$, $p < .001$). In addition, Communicators who engaged in Unconditional Honesty were seen as less likely to tell prosocial lies (M = 2.62, SD = 1.48) than those who Looked (M = 4.09, SD = 1.73; $t(446) = -9.71$, $p < .001$).

Relationships. Participants preferred Communicators who were Unconditionally Honest over Communicators who Looked in all relationships: as friends, coworkers, managers and leaders. See Figure 4 for statistical details.

Behavioral intentions. Participants were more likely to trust the advice of (consistent with Study 2), and ask for feedback from, Communicators who were Unconditionally Honest. See Figure 4 for statistical details.

Mediation. We conducted a mediation model with 10,000 samples using condition as the independent variable (1 = Unconditional Honesty, 0 = Looking), likelihood of telling harmful and prosocial lies as mediators, and judgments of morality as the dependent variable (PROCESS Macro for SPSS, Model 4; Hayes, 2013). Here, we find that the both the tendency to tell prosocial lies (indirect effect = 0.14, SE = 0.05, 95% CI [0.05, 0.24]) and the tendency to tell harmful lies (indirect effect = 0.28, SE = 0.06, 95% CI [0.17, 0.40]) mediate the relationship between condition and ratings of morality. Specifically, Communicators who engaged in Unconditional Honesty were judged as having a lower tendency to tell harmful lies (b = -0.79, p < .001), and the perceived tendency to tell harmful lies was negatively related to judgments of morality (b = -0.35, p < .001). Conversely, Communicators who engaged in Unconditional Honesty were judged as having a lower tendency to tell prosocial lies (b = -1.48, p < .001). The perceived tendency to tell prosocial lies was then negatively related to judgments of morality (b = -0.09, p = .002), albeit to a lesser extent that the perceived tendency to tell harmful lies was, consistent with Study 3.

Figure 4

Participants' Preferences for Relationships with and Behavioral Intentions towards Communicators who engaged in Unconditional Honesty vs. Looking

Notes: *One person did not provide a response for the question about trusting advice which is why the degrees of freedom is only 445. Error bars represent 95% confidence intervals around the mean.

Discussion

Study 4 extends our investigation by documenting additional downstream social consequences of moral preferences for unconditional honesty: people prefer unconditionally honest communicators as social partners across a range of relationships and are more likely to ask them for advice and feedback. Study 4 also provides further evidence that positive moral judgments of communicators who engage in unconditional honesty are driven – at least in part – by the belief that these communicators are less likely to tell harmful lies.

Study 5: Eliminating the Risks of Conditional Honesty

In Study 5, we test a final prediction of our account: if it is clear that communicators who engage in looking (i.e., conditional honesty) will not tell harmful lies, the reputational benefits of unconditional honesty should be attenuated. In Study 5, we use a similar paradigm to the one we used in Studies 2 and 3 and manipulate whether communicators who engage in looking can precommit to a specific course of action that prevents harmful lies. We predict that communicators who engage in looking and precommit to avoiding harmful lies will no longer be penalized for the decision to look. Specifically, we expect that communicators who look, but *only* allow themselves to tell prosocial lies, will be expected to tell fewer harmful lies overall, and therefore be judged as more moral.

Method

Participants. As preregistered, we aimed to recruit 800 participants using the Academic Prolific platform. We ended up with a final sample of 770 participants (M_{age} = 34.66, SD_{age} = 12.24; 384 men, 366 women, 15 prefer to self-describe, 5 prefer to not answer).

Procedure and materials. Participants were randomly assigned to a condition from a 2 (Decision Strategy: Unconditional Honesty vs. Looking) x 2 (Decision Option: Precommitment vs. No Precommitment) between-subjects design. The No Precommitment conditions were nearly identical to Unconditional Honesty and Looking conditions in Study 2, except for two key changes.

First, we changed the possible values of $A and $B. As in Study 2, participants knew that if they reported that the coin landed on heads, their partner would earn $A, and if they reported that the coin landed on tails, their partner would earn $B. In Study 5, however, Communicators knew that there was a 50% chance that $A was -$2 and $B was $5, and a 50% chance that $A was $5 and $B was -$2. We determined these possible values of $A and $B based on the results

of a small pilot study (N = 35) that examined the threshold at which people would endorse a prosocial lie over a harmful truth. Specifically, the pilot participants learned that a coin flip landed on heads and that honestly reporting this outcome would result in -$2 for another participant. Participants then responded to the question "Imagine the coin landed on Heads. If A = -$2.00, what value of $B would make it ethical to report tails, rather than heads?" In doing so, they indicated the minimum value that would make it more ethical to lie about the coin flip outcome than to tell the truth. All 35 pilot participants reported it would be more ethical to tell the prosocial lie if B was at least $5 when A was -$2. This pilot confirmed that we created a paradigm in which participants would indeed believe that prosocial lying was more ethical than harmful truth-telling (as in Study 1), which we did not explicitly test in Studies 2 and 3.

Second, we slightly changed the language in the Looking option. Specifically, the option read, "I'd like to base my decisions on the values of $A and $B." Participants knew that if the Communicator chose this option, they would see the values associated with $A and $B and then make their decision (reporting heads or tails). Other than these values and language modifications, the No Precommitment conditions matched Study 2. The Precommitment conditions, however, extends our investigation by adding the option to avoid harmful lies.

In the Precommitment condition, we modified the Looking options further. Participants learned that if the Communicator chose, "I'd like to base my decisions on the values of $A and $B," they would then precommit to decisions for each possible value of $A and $B. Specifically, Communicators would indicate what they would do if $A turned out to be $5 and $B turned out to be -$2, *and* what they would do if $A turned out to be -$2 and $B turned out to be $5. These decisions would then automatically be instituted by the computer once the true values of $A and $B were revealed. In our Precommitment/Looking condition, participants always learned that the

Communicator chose "I'd like to base my decisions on the values of $A and $B" and then precommitted to reporting whichever coin flip outcome (heads or tails) yielded a $5 bonus for their partner. This precommitment reflects conditional honesty, but it also eliminates the possibility of telling a harmful lie: the Communicator precommitted to telling the truth if doing so was helpful to their partner and to lying if doing so was helpful to their partner. Effectively, precommitment allowed Communicators to signal that they would avoid telling harmful lies. Appendix A, Figure A2 depicts the exact text participants saw in the Precommitment/Looking condition.

Participants had to pass comprehension checks to confirm that they both understood the Decision Strategy and the Decision Options of the Communicator. Participants who answered a question incorrectly had the chance to review the instructions and answer the questions again. Participants who answered any questions incorrectly on the second try were automatically kicked out of the survey.[4]

Dependent Variables.

Morality. Our primary dependent variable was the judgments of the Communicator's morality using the same three item composite used in Studies 2, 3, and 4 ($\alpha = .923$).

Tendency to tell Harmful and Prosocial Lies. Participants reported their belief that the Communicator would tell harmful and prosocial lies in general, using the same items as Study 3.[5]

[4] We note that the nature of the comprehension checks did lead to differential attrition across conditions. Participants were more likely to fail the comprehension checks when they were in the Precommitment condition, leading to somewhat uneven cell sizes. Therefore, it is possible that remaining participants in the Precommitment condition were more thoughtful, attentive participants than those in the No precommitment condition.

[5] We also measured two alternative mechanisms - the extent to which participants believed that Communicators who Looked would go down a "slippery slope" of deceptive behavior in the

Results

Morality. A two-way ANOVA on judgments of morality revealed a main effect of Decision Strategy ($F(1,766) = 18.96$, $p < .001$); Communicators who engaged in Unconditional Honesty were viewed as more moral than those who engaged in Looking. There was also a main effect of Decision Option ($F(1,766) = 11.83$, $p = .001$) such that Communicators who expressed precommitment were viewed as more moral than those who did not precommit. All descriptive statistics for Study 5 are presented in Table 1.

Importantly, these results were qualified by a significant interaction of Decision Strategy and Decision Option ($F(1,766) = 18.15$, $p < .001$). In the No Precommitment condition, we replicated our previous findings: Communicators who engaged in Unconditional Honesty were judged as more moral than those who engaged in Looking ($t(416) = 6.71$, $p < .001$). However, in the Precommitment condition, Communicators who engaged in Unconditional Honesty were judged as equally moral as Communicators who expressed prosocial precommitment before Looking ($t(350) = .060$, $p = .952$). Figure 5 depicts these results.

Table 1

Descriptive Statistics in Study 5

	No Precommitment	Precommitment

future (Anderson et al., 2023; Garrett, Lazzaro, Ariely & Sharot, 2016), and that they had poor moral standards broadly (Huppert et al., 2023). Although beliefs that the communicator internalizes the value of honesty do mediate the effect of Unconditional Honesty (vs. Looking) on moral judgments, this was the case regardless of whether the communicator precommitted to avoiding harmful lies. These results suggest that concerns about overall moral standards may be independent of concerns about the risks of harmful lies. We find no evidence of mediation through slippery slope beliefs. When we run moderated mediation models that include these two additional alternative mechanisms, we also find significant evidence of moderated mediation through prosocial lying, which is not the case in the analyses reported in the main manuscript. See SOM 1.5 for more details.

Variable	Unconditional Honesty	Looking	Unconditional Honesty	Looking
Morality	5.41 (1.17)	4.69 (1.03)	5.34 (1.12)	5.34 (1.35)
Tendency to tell Harmful Lies	2.33 (1.25)	3.08 (1.16)	2.53 (1.23)	2.19 (1.10)
Tendency to tell Prosocial Lies	3.14 (1.41)	4.32 (1.22)	3.30 (1.51)	5.38 (1.45)

Notes. Standard deviations appear in parentheses.

Figure 5
Moral Judgment of the Communicator in Study 5

Notes. Error bars represent 95% confidence intervals around the mean. **signifies $p < .001$.

Tendency to tell Harmful Lies. Judgments of harmful lies mirrored moral character judgments. A two-way ANOVA on beliefs about the Communicator's tendency to tell harmful lies revealed a main effect of Decision Strategy ($F(1,766) = 5.84$, $p = .016$). As in Studies 3 and 4, Unconditional Honesty signaled that Communicators would be less likely to tell harmful lies compared to Communicators who Looked. This analysis also revealed a main effect of Decision Option ($F(1,766) = 16.07$, $p < .001$). Communicators who expressed precommitment were rated

as less likely to tell harmful lies than those who did not precommit, consistent with the nature of the manipulation.

These effects were qualified by a significant interaction of Decision Strategy and Decision Option ($F(1,766) = 38.63$, $p < .001$). In the No Precommitment condition, Communicators who engaged in Unconditional Honesty were judged as less likely to tell harmful lies than Communicators who Looked at the consequences ($t(416) = -6.29$, $p < .001$). However, in the Precommitment condition, Communicators who engaged in Unconditional Honesty were judged as *more* likely to tell harmful lies than those who Looked ($t(350) = 2.62$, $p = .009$).

Tendency to tell Prosocial Lies. A two-way ANOVA on beliefs about the Communicator's tendency to tell prosocial lies revealed a main effect of Decision Strategy ($F(1,766) = 258.06$, $p < .001$). Unconditionally Honest Communicators were rated as less likely to tell prosocial lies than Communicators who Looked. We also found a main effect of Decision Option ($F(1,766) = 35.66$, $p < .001$) such that Communicators who expressed precommitment were rated as more likely to tell prosocial lies than those who did not precommit, consistent with the nature of the manipulation.

Notably, these effects were qualified by a significant interaction of Decision Strategy and Decision Option ($F(1,766) = 19.77$, $p < .001$). In the No Precommitment condition, Communicators who Looked were judged as more likely to tell prosocial lies than Communicators who were Unconditionally Honest ($t(416) = -9.12$, $p < .001$). However, this effect was significantly larger in the Precommitment condition in which Communicators who Looked were rated as much more likely to tell prosocial lies than Unconditionally Honest Communicators ($t(350) = -13.07$, $p < .001$).

Moderated Mediation Model. We conducted a moderated mediation model with 10,000 samples using Decision Strategy as the independent variable (1 = Unconditional Honesty, 0 = Looking), Decision Option as the moderator, and judgments of the Communicator's morality as the dependent variable (PROCESS Macro for SPSS, Model 7; Hayes, 2013). We conducted this analysis using tendency to tell harmful lies and tendency to tell prosocial lies as simultaneous mediators.

Results of the moderated mediation analysis are presented in Table 3. Consistent with our theorizing, we found significant moderated mediation through beliefs about the Communicator's tendency to tell harmful lies (index of moderated mediation = -0.55, SE = 0.09, 95% CI [-0.73, -0.37]). Engaging in Unconditional Honesty, relative to Looking, led to lower concerns about harmful lies (b = -0.75, p < .001) when Communicators who Looked did not express precommitment, but led to higher concerns about harmful lies (b = .33, p < .05) when Communicators who Looked precommitted to making a prosocial decision. Concerns about harmful lies were negatively correlated with judgments of morality (b = -0.51, p < .001; indirect effect in No Precomittment Condition = 0.38, SE = 0.07, 95% CI [0.26, 0.51]; indirect effect in Precomittment Condition = -0.17, SE = 0.06, 95% CI [-0.29, -0.05]). We did not detect significant moderated mediation through beliefs about the Communicator's tendency to tell prosocial lies (index of moderated mediation = -0.04, SE = 0.03, 95% CI [-0.10, 0.01]). Although Unconditional Honesty lowered concerns about prosocial lying in both the No Precommitment (b = -1.18, p < .001) and Precommitment (b = -2.09, p < .001) conditions, prosocial lying was not significantly associated with moral judgment in this study (b = .05, p = .068; indirect effect in No Precomittment Condition = -0.05, SE = 0.03, 95% CI [-0.12, 0.01]; indirect effect in Precomittment Condition = -0.10, SE = 0.06, 95% CI [-0.22, 0.02]).

Table 3

Moderated Mediation Analyses from Study 5

Mediators	No Precommitment	Precommitment	Index of Moderated Mediation
Tendency to tell harmful lies	**0.26, 0.51**	**-0.29, -0.05**	**-0.73, -0.37**
Tendency to tell prosocial lies	-0.12, 0.01	-0.22, 0.03	-0.10, 0.01

Notes. Each set of numbers indicates the lower-level and upper-level 95% confidence intervals around the indirect effect of the corresponding mediator. Bold typeface indicates significant effects (i.e., confidence intervals do not contain zero).

Discussion

In Study 5, we manipulated whether looking could lead to harmful lies. When we eliminated the possibility that communicators who look will tell harmful lies, the reputational benefits of unconditional honesty were attenuated. These results are consistent with our account: when conditional rule-following no longer signals a communicator's likelihood of telling harmful lies, unconditional honesty is no longer reputationally advantageous. We replicate these results in Studies S7 and S9 (see SOM 2.7 and 2.9), in which we also examine judgments of unconditional honesty and looking after the outcomes of these decision strategies are realized (i.e., after participants learn whether they lead to prosocial or harmful truths or lies). In these studies, looking was seen as less ethical than unconditional honesty when both strategies resulted in harmful, but not prosocial, outcomes.

In line with Studies 3 and 4, our mediation results also suggest that looking (conditional honesty) is troubling not simply because it could lead to harmful lies in the present context, but rather because it signals that a person may be likely to tell those lies in the future. Communicators who looked for information were judged as more likely to tell harmful lies in

general compared to communicators who were unconditionally honest, which led them to be seen as less ethical.

General Discussion

Across our studies, we find evidence that people judge unconditional honesty as a normatively moral decision strategy and reward communicators who engage in it. Though people believe that some lies – namely, prosocial lies – are ethical, it is not always possible to know the outcomes associated with honesty ahead of time. When the outcomes of honest behavior are unknown, people value unconditional honesty because it prevents the most unethical actions – namely, harmful lies – from occurring, even though unconditional honesty also eliminates the possibility of telling prosocial lies.

In Study 1 (as well as Study S1 in the SOM), we establish the moral preference for unconditional honesty. We find that people believe that communicators are obligated to condition their communication decisions on the social consequences of honesty (and tell prosocial lies rather than harmful truths) when the social consequences of honesty are known. However, people also believe that communicators should avoid learning about these consequences if they are initially uncertain about them, upholding a policy of unconditional honesty.

In Study 2, we begin to explore the reputational consequences of this preference. We find that communicators who engage in unconditional honesty are judged to be more moral, and are trusted more as advisors, than communicators who look at the social consequences of honesty before deciding whether to tell the truth. In other words, unconditional honesty is rewarded, despite reflecting strategic ignorance of social welfare. In Studies 3-5, we directly test our proposed mechanism that engaging in unconditional honesty is rewarded because it signals

that the communicator will not engage in harmful lies. In Studies 3 and 4, we test this through mediation, using two different paradigms. In Study 4, we also explore how unconditional honesty influences social preferences, finding that communicators who engage in unconditional honesty are preferred across a variety of social relationships. In Study 5, we test our theoretical account through moderation. We find that communicators who engage in unconditional (versus conditional) honesty are no longer seen as more moral if it is certain that conditional honesty will not lead to harmful lies.

Supplementary studies and meta-analytic results

A number of supplementary studies provide further evidence of the moral preference for unconditional honesty, rule out alternative explanations, and begin to examine the boundaries of our effects. We report these studies in our supplementary online materials. Here, we provide a meta-analysis of these studies and review a few of their key results. We encourage interested readers to examine our supplement for further details.

Our internal meta-analysis included 13 studies (Studies 2-5 from the main manuscript, and Studies S2-S10 reported in the online supplement). In each study, we included only the conditions that examined judgments of Unconditional Honesty and Looking (omitting conditions in our supplemental studies that that examined Unconditional Lying or Unconditional Prosociality; see SOM for details). We used the meta-analysis package of SPSS 28, using random-effects modeling and inverse variance weights to assign weights to studies based on their sample size (Lipsey & Wilson, 2001). The summary statistics from each condition of each study are provided in Appendix B (see Table A1 for statistics and Figure A3 for a forest-plot). Overall, we find consistent evidence that Unconditional Honesty is judged to be more ethical than Looking; meta-analytic d = .465, 95% CI [0.32, 0.61]. Though we find a small-medium meta-

analytic effect size, we note that our meta-analysis includes conditions in which we expected (and did find) an attenuation of our effects (e.g., in Study 5: Precommitment when Unconditional Honesty and Looking both prevented harmful lies). Therefore, this analysis suggests that in general, Unconditional Honesty is likely to be seen as more moral than Looking. However, for completeness, we also conducted sub-group analyses to examine the overall effect of Unconditional Honesty versus Looking when Looking could, versus could not, yield harmful outcomes. When Looking could lead to harm, we continue to see a robust moral preference for Unconditional Honesty; meta-analytic d = .534, 95% CI [0.39, 0.67]. However, when Looking could not lead to harm, Unconditional Honesty was not seen as more moral than Looking; meta-analytic d = -.115, 95% CI [-0.30, 0.06]. These results are consistent with our test of precommitment in Study 5, as well as our broader theory, and highlight the moral importance of avoiding harmful lies.

Notably, these studies also address a number of important questions about the nature of people's preference for unconditional honesty. In Studies S2, S3, and S5, for example, we explore whether people have a greater preference for unconditional honesty when honesty could cause personal harm to the communicator, rather than social harm. Interestingly, we find little evidence that the preference for unconditional honesty stems from concerns about the communicator's selfishness. In Studies S7 and S9, we examine judgments of unconditional honesty, after the consequences of these decisions are known. Whereas the studies in the main manuscript demonstrate that unconditional honesty is rewarded a priori (before the outcomes are realized), these studies find that unconditional honesty is also rewarded post hoc, after it causes harm. Intentional ignorance of honesty's consequences seems to be rewarded, even once the social harm of that ignorance is realized.

Another set of studies (namely Studies S6 and S7) began to explore the interesting question of whether the preference for unconditional rule-following extends beyond the domain of honesty. In these studies, we manipulate whether communicators start with information about what is honest (and have to decide whether to find out if honesty would help or harm others before making a communication decision, as in Studies 2-5) or with information about what is helpful (and have to decide whether to find out if helping others entails lying or truth-telling before making a helping decision). In doing so, we test whether people have equivalent preferences for unconditional honesty and unconditional prosociality. We find that the rewards of unconditional rule-following apply to the domain of honesty but not to prosociality, suggesting that honesty may be somewhat unique.

Contributions

Understanding honesty. Honesty is one of the most important values to everyday moral decision-making. Yet, our attitudes towards honesty are not easily explained by existing theories of moral judgment. Existing work on deontological intuitions, for example, has primarily been developed to understand the psychology of harm aversion, which as others have noted, may be largely unrelated to the psychology of honesty (e.g., Kahane et al., 2012, 2018). Our work is the first that we know of to develop a theory of honesty that accommodates the preference for unconditional honesty *and* the preference for certain lies.

Therefore, this work helps us to understand seemingly hypocritical attitudes towards honesty. People routinely lie in everyday life and recognize that many common lies are prosocial (DePaulo et al., 1996). If people endorse prosocial lies, then we might expect education and communication to reflect these beliefs. For example, people might state that honesty is *often* the best policy but lying is sometimes okay. However, people – including parents, teachers, and

public leaders – instead tend to invoke absolute language around honesty, endorsing the idea that honesty is *always* the best policy. Thus, there is an inconsistency in people's messaging around honesty and actual honest behavior (Huppert et al., 2023). Our current studies add insight into why such inconsistency may be valued in the context of honesty, in contrast to many other domains in which inconsistencies lead to negative judgments (Effron et al., 2018; Jordan et al., 2017). Absolute rules prohibiting lying are valued because they prevent the most immoral actions when the consequences of honesty are uncertain: harmful lies.

Preferences for unconditional rule following. This work also helps us understand the preference for absolute moral rules broadly, in the context of ethical dilemmas. A growing body of research on the benefits of "uncalculating cooperation" (Capraro & Kuilder, 2016; Hoffman et al., 2015; Jordan et al., 2016) finds that actors who cooperate without first finding out the personal consequences of doing so are judged as more trustworthy than actors who find out the personal consequences of cooperation before making cooperation decisions. Similarly, taking less time to make a moral decision (Critcher et al., 2013; Jordan et al., 2016) or relying on emotion rather than reason (Barasch et al., 2014; Levine et al., 2018) is perceived to signal morality and trustworthiness, in part, because making intuitive decisions conveys a genuine commitment to moral behavior (Capraro & Kuilder, 2016; Evans & Van de Calseyde, 2017; Van de Calseyde et al., 2014). Scholars have suggested that this body of work can explain why people appreciate unconditional morality generally, including the decision to engage in unconditional honesty. These effects, however, have only been tested in right-wrong situations (i.e., situations in which there are potential conflicts between following a moral principle and pursuing one's self-interest). In the uncalculating cooperation paradigm, for example, an actor can cooperate (the moral, prosocial choice) without looking at how costly cooperation would be for them, or an

actor can first look to see how costly cooperation is, which might tempt them to defect (the immoral, selfish choice; Jordan et al., 2016). In this paradigm, it is assumed that an actor would only "look" if they were considering acting immorally, which presumably signals poor character. In contrast, uncalculating cooperation signals commitment to cooperation, suggesting that a person is not conflicted between helping others and acting selfishly.

Beliefs about decision conflict, however, cannot easily explain the preference for moral rule-following in dilemmatic contexts. People experience, and are expected to experience, high decision conflict when faced with ethical dilemmas (i.e., situations in which there is a conflict between two moral principles; Kidder, 1995; Kohlberg, 1971; Zhang et al., 2018). Furthermore, as discussed, in the context of honesty-prosociality dilemmas, people often value dishonesty when they know that honesty will cause harm. Therefore, existing theories do not provide a clear explanation for why people would reward unconditional honesty, even when honesty is associated with harm. The present theory does. In right-right (i.e., dilemmatic) contexts, following a moral rule prevents the most unethical actions from occurring. Interestingly, our theory predicts that unconditional honesty is seen as moral even though it reflects strategic ignorance of the social harm caused by one's communication, which also expands our understanding of how information acquisition influences moral judgment.

Normative and descriptive theories of moral judgment. Furthermore, our account helps explain and bridge conflicting evidence on which normative theories best describe lay moral judgment, thereby providing insight into the perceived purpose of moral rules. Some scholars have suggested that lay people are best described as intuitive deontologists, who place importance on following universal moral rules (e.g., Everett et al., 2016; Greene, 2007; Jordan et al., 2016; Greene et al., 2001). In contrast, other scholars have suggested that lay people are best

described as intuitive virtue-ethicists, who place importance on deducing an actor's internal character based on how the actor weighs conflicting moral concerns in a given situation (e.g., Critcher et al., 2020; Landy & Uhlman, 2018). Our account suggests that both of these normative lenses characterize lay judgments of honesty, albeit in different contexts.

When evaluating individual actions in a given situation with known consequences, people care about reducing harm. People believe that the duty to be honest should be constrained by concerns of harm, consistent with a virtue-ethics lens. However, when facing situations with unknown moral outcomes, people believe that it is more moral to follow a categorical policy of unconditional honesty than to seek out information that would put them in situations where moral rules might be compromised. This belief is consistent with a deontological lens.

Although these beliefs may seem contradictory at first, both are driven by an overall desire to avoid harm and minimize unethical outcomes. In this way, both the belief that some lies are acceptable and the categorical prohibition of dishonesty are compatible with utilitarian calculations of harm (consistent with notions of 'rule' utilitarianism; Kahane et al., 2018; Scheffler, 1982; Sen, 1983 and moderate deontology; Holyoak & Powell, 2016; see also Ditto & Liu, 2011; Liu & Ditto, 2013). However, the different decision contexts – whether people are evaluating an action with known consequences or a policy that applies to future decisions with unknown consequences – seem to shift people's attention from what actions minimize harm in the given situation (a prosocial lie over a harmful truth) versus what actions minimize harm overall, across situations (avoiding harmful lies in general). Taken together, our work highlights how the psychology of honesty hinges on multiple perspectives of moral judgment.

Roadmap for future research

Understanding the communicator perspective

We encourage future research to explore whether communicators anticipate others' preference for unconditional honesty. Are communicators attuned to the social rewards associated with unconditional honesty? If so, does this lead them to engage in unconditional honesty? Our lived experiences with honesty and dishonesty make this seem unlikely. However, recent lab studies that used a paradigm similar to the one in the present paper found that communicators often choose not to find out whether their honesty causes harm (Levine & Munguia Gomez, 2021). In three of our supplemental studies (see SOM 3 for a review), we also found that the majority of communicators chose to be unconditionally honest.

However, in one study of communicators in which looking was the default (see SOM 1.3; also reported in Study 3), we found that the majority chose to look at the consequences of honesty. Future work should examine communicators' decisions more thoroughly. It is possible that communicators' decisions involving honesty are influenced not only by the defaults presented to them, but also by the degree to which they focus on the reputational benefits accorded by observers, or the potential relational costs accorded by the targets of harmful truths. Manipulating whether or not communicators are aware that targets or external observers are judging their decision strategy would be helpful for understanding how communication decisions are influenced by expected reputational benefits.

Studying different types of lies

Future work would also benefit from exploring how judgments of unconditional honesty change when people think about different types of lies. We characterize prosocial lies as those that benefit the recipient of the lie in the long-run, and harmful lies as those that hurt the recipient of the lie in the long-run. However, harmful lies can stem from prosocial or malevolent *motives*. For example, in Studies 1 and 4, we operationalize harmful lies as lies that would

undermine the well-being of the recipient in the long-run but could be motivated by prosocial motives such as compassion, politeness, or conflict-avoidance. In these studies, the communicator could offer candid criticism or false praise to a colleague about their suit (Study 1) or their report (Study 4). False praise is harmful to the colleague if it prevents them from changing and improving something they had control over, even if it also spares the colleague from emotional harm. In our remaining studies, we operationalized harmful lies as lies that lead to monetary harm for recipients. These lies are less likely to be attributed to prosocial motives. Indeed, we find that the attributions associated with harmful lies were more cynical in Study 3 (which featured monetary harm) versus Study 1 (see SOM 1.1.2.2 and SOM 1.3.1 for details).

Importantly, we find that people still reward unconditionally honest communicators, regardless of a communicator's potential motives for telling a harmful lie. However, it is possible that when harmful lies are relatively innocuous or prosocial lies are extraordinarily helpful (as in the classic example of lying to a murderer at your door), preferences for unconditional honesty may be attenuated. In a few of our supplemental studies, we explore the nuances around telling prosocial lies, finding that a perceived tendency to tell prosocial lies is still negatively associated with judgments of morality, albeit to a lesser extent than the perceived tendency to tell harmful lies (see SOM 2.3 and 2.4). To understand the complexity of honesty in everyday life, it is important to further explore how judgments of unconditional honesty relate to different motives for and consequences of lying.

Addressing Constraints on Generality

Altogether, our work demonstrates that people have a robust preference for unconditional honesty, stemming from the desire to avoid harmful lies. We expect these results to replicate when similar experimental designs are employed within similar populations. However, below we

discuss three constraints of our methodological approach, and potential future research to address them, in more detail.

Unconditional honesty in everyday conversation. The use of tightly controlled experiments allowed us to examine mechanisms underlying preferences for unconditional honesty. However, future work should examine preferences for unconditional honesty in everyday, real-world situations. Across all of our studies, unconditional honesty and looking were operationalized as distinct, mutually exclusive decision strategies. In natural conversation, it may be possible to integrate these approaches in order to hedge against the relational costs of harmful truths. For example, consider the case of a friend who asks how an ill-fitting outfit looks (as in Study 1). While a communicator could immediately respond by stating the truth (Unconditional Honesty) or by seeking out information about whether the friend could change (Look), they may also use a more integrative strategy. For example, they may choose to start with a prosocial lie – responding that their friend looks great – and *then* ask for more information (i.e., casually asking if the friend has any other outfits they were considering wearing). If a communicator learns that their friend cannot change, the decision tree ends benevolently at a prosocial lie rather than a harmful truth that would have resulted from unconditional honesty. On the other hand, if the friend can change, the communicator can follow their prosocial lie with a prosocial truth by recommending that a different outfit would be even better. This strategy, "Prosociality *plus* Looking," prioritizes benevolent communication by maximizing the welfare of targets across cases. Importantly, this strategy also ensures that harmful lies are not told.

While this strategy avoids an explicitly harmful lie, it is also possible that asking for information in and of itself conveys some of the harmful truth. When a communicator asks their friend whether they have other outfits to wear, this may signal that the communicator does not

like the current outfit very much. In other words, looking for more information conveys that the communicator is withholding a harmful truth. In a post-test of Study 1 (see SOM 1.1.2), we find partial support for this idea. When asked to what extent the target could infer the communicator's true opinion from their looking behavior, participants reported that the target *could* likely infer the communicator had a negative opinion of the target's suit (significantly higher than the midpoint on a 7-point Likert scale, $M = 4.33$, $SD = 1.36$, $p < .05$). However, this inference did not predict how moral looking was perceived to be ($p = .868$), suggesting that this concern cannot explain preferences for unconditional honesty overall. Future research should investigate whether people employ hybrid approaches in naturalistic settings and, relevant to the present investigation, how these approaches are judged.

Moral rule-following across domains and cultures. Future research also ought to examine how worst outcome avoidance (Zlatev et al., 2020) applies to other moral domains. In two supplemental studies (Studies S6 and S7), we found that although people judged unconditional honesty to be more ethical than looking, people did not judge unconditional prosociality in the same way. However, more research is needed to understand why, given that both unconditional prosociality and unconditional honesty prevent harmful lies. It is possible that seeking out information about the *consequences* of a moral behavior signals greater moral flexibility than seeking out information about the presence of any moral conflict (Kreps & Monin, 2014). As a result, seeking out information about consequences may activate concerns about negative moral outcomes to a greater extent.

The preference for unconditional rule-following may also be more or less prominent in different cultures. People in Eastern cultures tend to have a more holistic thought style, whereas people in Western cultures tend to have a more analytic thought style (Nisbett et al., 2001).

These differences in cognition may have implications for how people reason about moral dilemmas under uncertainty. In Western cultures, an emphasis is placed on principles and logic, so we might expect that an analytic thought style leads to a preference for absolute, categorical rules. In contrast, cultures that employ a holistic thought style or place greater weight on relational obligations may actually value those who choose to look at the consequences of honesty first because it resolves uncertainty about how a course of action affects one's interpersonal relationships. Furthermore, the unethicality of lying itself may differ across cultures. There is evidence to suggest that a culture's levels of individualism-collectivism can impact judgments of deception (Tong et al., 2023). For example, children from Chinese cultures believe that lying is more acceptable than Canadian children, particularly lying aimed at downplaying individual achievements, protecting collective interests, or promoting social cohesion (Lee at al., 2001; Fu et al., 2007). Similarly, as adults, Chinese participants believe lying as a form of modesty or when parenting is more acceptable than American or Canadian participants (Heyman et al., 2013; Fu et al., 2001). Therefore, whether or not lies are seen as harmful may differ across cultural contexts and, consequently, influence preferences for unconditional honesty.

Conclusion

People simultaneously claim that honesty is the best policy and that prosocial lying is ethical. The present research develops and tests a theory to explain these seemingly incompatible beliefs. People endorse unconditional honesty as a policy when the social consequences of honesty are unknown because it mitigates the risk of harmful lies. Avoiding the worst moral outcome (harmful lies) is seen as more important than optimizing every moral decision. This account helps explain the psychology of honesty, as well as moral rules more broadly.

Appendix A: Stimuli Used in Studies 4 and 5

Figure A1.

Stimuli Used in the Unconditional Honesty (top) and Looking Conditions (bottom) of Study 4

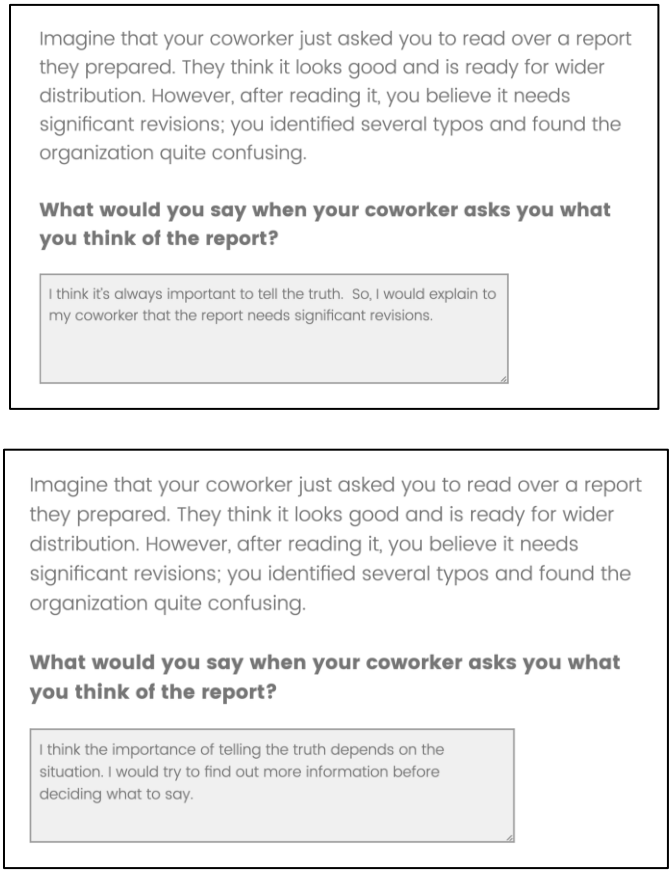
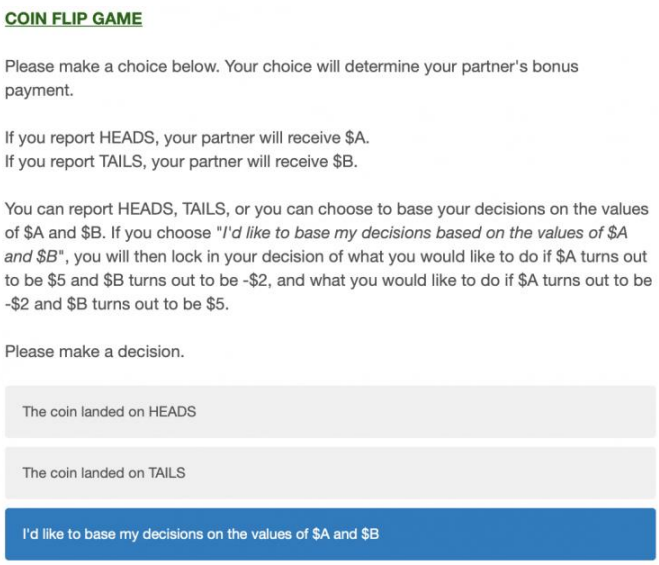


Figure A2.

Stimuli Depicting Communicator Choices in the Precommitment/Looking Condition of Study 5

First, participants learned about the Coin Flip Game and saw that the coin had landed on HEADS. Then, participants saw the Communicator's initial choice. The Communicator below selected the Looking option (indicated by blue highlighting).



If the Communicator selected the Looking option, participants saw the Communicator's precommitment decisions on the next page. The Communicator always precommitted to the prosocial choices. If the Communicator chose Unconditional Honesty ("The coin landed on HEADS") on the first screen, no second page was presented.

Appendix B: Effect Sizes and Meta-analytic Results

Table A1.

Meta-analysis of all Studies that Examine Moral Judgments of Unconditional Honesty and Looking

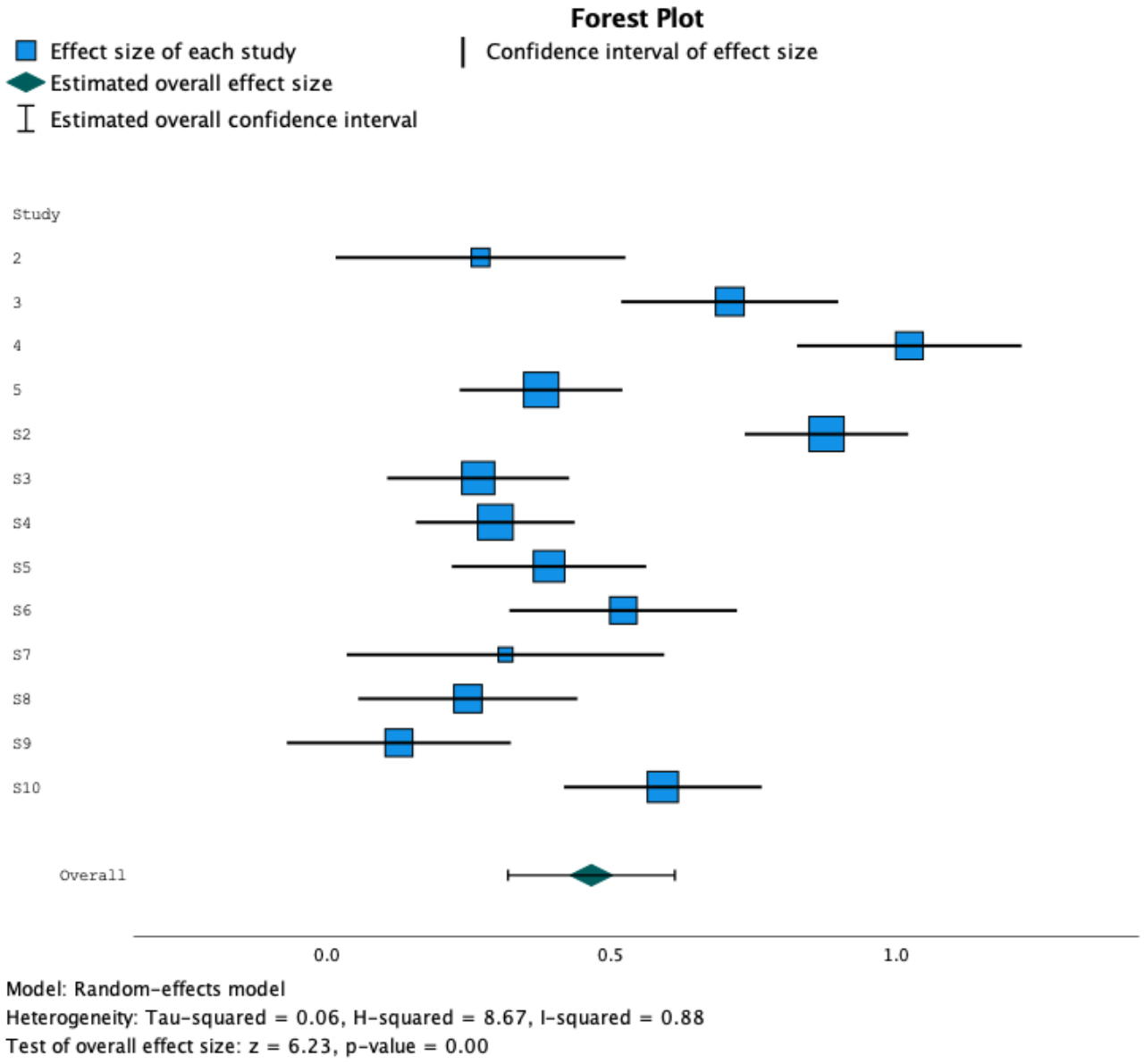
			Unconditional Honesty			Looking				95% CI	
Study	**Main finding**	**Additional factors**	***M***	***SD***	***n***	***M***	***SD***	***n***	***d***	***Lower***	***Upper***
2	Unconditional honesty is seen as more moral than Looking	N/A[a]	4.37	1.14	114	4.06	1.15	126	0.273	0.019	0.528
3	Unconditional honesty is seen as more moral than Looking and signals a lower propensity to tell harmful lies	N/A[a]	5.48	1.22	226	4.68	1.03	224	0.706	0.515	0.896
4	Unconditionally honest communicators are seen as more moral and as better relationship partners	N/A[a]	6.00	0.86	224	4.89	1.27	224	1.03	0.829	1.223
5	Unconditional honesty is seen as more moral than Looking, unless Looking is paired with precommitment to avoid harm	No precommitment[a]	5.41	1.17	168	4.69	1.03	250	0.669	0.468	0.869
		Precommitment [b]	5.34	1.12	198	5.34	1.35	154	0.007	-0.204	0.217
		Overall	5.38	1.14	366	4.93	1.20	404	0.377	0.234	0.520
S2	Unconditional honesty is seen as more moral than Looking, regardless of whether these decision strategies influence payoffs for the self or others	Personal consequences	5.32	1.13	221	4.22	1.25	216	0.924	0.727	1.121
		Social consequences	5.27	1.11	187	4.35	1.10	193	0.834	0.624	1.043
		Overall[a]	5.29	1.12	408	4.28	1.18	409	0.883	0.739	1.026
S3	Unconditional honesty is seen as more moral than Looking, regardless of whether these decision strategies influence payoffs for the self or others	Personal consequences	4.51	1.05	101	3.93	1.56	104	0.437	0.159	0.713
		Social consequences	4.13	1.15	99	3.94	1.35	101	0.149	-0.129	0.426
		Personal & social cons.	4.38	1.31	97	4.01	1.92	107	0.224	-0.052	0.499
		Overall[a]	4.34	1.18	297	3.96	1.63	312	0.266	0.107	0.426

S4	Unconditional honesty is seen as more moral than Looking, regardless of how salient harm is	Harm less salient	4.63	1.12	192	4.28	1.23	202	0.301	0.102	0.500
		Harm salient	4.73	1.03	199	4.42	1.11	207	0.292	0.096	0.488
		Overall[a]	4.68	1.07	391	4.35	1.17	409	0.297	0.157	0.436
S5	Unconditional honesty is seen as more moral than Looking, regardless of whether a communicator also engages in Unconditional when these strategies influence payoffs for the self	HWOL, personal cons.	5.45	1.07	86	4.74	1.25	89	0.612	0.308	0.915
		Looking, personal cons.	4.65	1.17	91	4.40	1.24	94	0.209	-0.081	0.497
		No information	5.12	1.07	85	4.68	1.11	94	0.398	0.101	0.694
		Overall[a]	5.06	1.15	262	4.60	1.20	277	0.390	0.220	0.561
S6*	Unconditional honesty is seen as more moral than Looking	N/A[a]	4.80	1.06	199	4.19	1.27	200	0.524	0.324	0.724
S7*	Unconditional honesty is seen as more moral than Looking when the strategies lead to social harm, but not when the strategies avoid harm	Prosocial outcome: Help[b]	4.82	0.88	49	4.99	1.03	53	-0.178	-0.567	0.211
		Prosocial outcome: Harm[a]	4.76	1.04	46	3.81	1.55	53	-0.710	-1.115	-0.301
		Overall	4.79	0.96	95	4.40	1.43	106	0.317	0.038	0.595
S8	Unconditional honesty is seen as more moral than Looking	N/A[a]	4.96	1.05	207	4.67	1.27	211	0.250	0.057	0.442
S9*	Unconditional honesty is seen as more moral than Looking when the strategies lead to social harm, but not when the strategies avoid harm	Prosocial outcome: Help[b]	5.24	1.30	101	5.56	1.28	97	-0.248	-0.527	0.033
		Prosocial outcome: Harm[a]	4.78	1.61	101	4.06	1.98	100	0.398	0.119	0.677
		Overall	5.01	1.48	202	4.80	1.83	197	0.127	-0.069	0.324
S10*	Unconditional honesty is seen as more moral than Looking, regardless of whether Looking is paired with prosocial intentions	No intentions expressed	5.47	1.06	196	4.89	0.99	206	0.565	0.366	0.764
		Prosocial intentions	*na*	*na*	*na*	4.86	0.94	199	0.61	0.408	0.812
		Overall[a]	5.47	1.06	196	4.88	0.97	405	0.597	0.423	0.770
								Overall meta-analytic estimate:	**0.465**	**0.319**	**0.612**
								Meta-analytic estimate when Looking could lead to harm[a]:	**0.534**	**0.394**	**0.674**
								Meta-analytic estimate when Looking did not lead to harm:	**-0.115**	**-0.295**	**0.064**

Notes. *indicates that the study also included additional cells (manipulations) that did not explore judgments of Unconditional Honesty. In Study S6, we only report results from the Context: Honesty conditions. Study S6 also explored judgments of Unconditional Prosociality (Context: Prosociality). In Study S7, we only report results from the Honesty Outcome: Truth and Context: Honesty conditions. Study S7 also explored judgments of Lying (Honesty Outcome: Lie) and Unconditional Prosociality (Context: Prosociality). In Study S9, we only report results from the Honesty Outcome: Truth conditions. Study S9 also explored judgments of Lying (Honesty Outcome: Lie). Full results, including all manipulations, are available in SOM 1 & 2. In Study S10, we include all conditions, but compare Unconditional Honesty to Looking both with and without intentions expressed. [a]indicates that the condition was included in the "Looking could lead to harm" subgroup analysis. [b]indicates that the condition was included in the "Looking did lead not lead to harm" subgroup analysis.

Figure A3.

Forest plot of the meta-analytic results of all studies that examine moral judgments of Unconditional Honesty and Looking



Context of the Research

A growing body of research shows that people reward decision-makers who are uncalculating and unconflicted when making ethical decisions. However, in the presence of ethical dilemmas – situations in which two ethical principles conflict – we might expect people to reward decision-makers who grapple with ethical tradeoffs, over those who simply prioritize one ethical principle over another. Indeed, the current author team expected this to be the case for ethical dilemmas involving honesty. Across a number of papers, Emma E. Levine has found that people do not typically hold unconditional stances on honesty – for example, they believe lying is ethical when it prevents unnecessary harm. Therefore, it seemed reasonable to predict that people would reward decision-makers who sought to find out if honesty caused unnecessary harm more than decision-makers who were always honest. Our author team was quite surprised when we found the opposite in initial studies! People reward decision-makers who are unconditionally honest, *despite* believing lying is sometimes ethical. This paper reflects our efforts to understand this puzzle. Emma E. Levine introduced the initial idea, and Sarah L. Jensen led the initial studies as part of her master's thesis. Mike W. White and Elizabeth Huppert joined the project due to overlapping interests in morality and honesty.

References

Abel, J. E., Vani, P., Abi-Esber, N., Blunden, H., & Schroeder, J. (2022). Kindness in Short Supply: Evidence for Inadequate Prosocial

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

Abi-Esber, N., Abel, J. E., Schroeder, J., & Gino, F. (2022). "Just letting you know…" Underestimating others' desire for constructive feedback. *Journal of Personality and Social Psychology*.

Anderson, R. A., Ruisch, B. C., & Pizarro, D. A. (2023). On the highway to hell: Slippery slope perceptions in judgments of moral character. *Personality and Social Psychology Bulletin*, 01461672221143022.

Backbier, E., Hoogstraten, J., & Terwogt-Kouwenhoven, K. M. (1997). Situational determinants of the acceptability of telling lies. *Journal of Applied Social Psychology*, 27(12), 1048-1062.

Barasch, A., Levine, E. E., Berman, J. Z., & Small, D. A. (2014). Selfish or selfless? On the signal value of emotion in altruistic behavior. *Journal of Personality and Social Psychology, 107*(3), 393-413.

Ben-Ner, A., & Halldorsson, F. (2010). Trusting and trustworthiness: What are they, how to measure them, and what affects them. *Journal of Economic Psychology*, 31(1), 64-79.

Berman, J. Z., & Silver, I. (2022). Prosocial behavior and reputation: When does doing good lead to looking good?. Current opinion in psychology, 43, 102-107.

Brown, P., Levinson, S. C., & Levinson, S. C. (1987). *Politeness: Some universals in language usage* (Vol. 4). Cambridge university press.

Capraro, V., & Kuilder, J. (2016). To know or not to know? Looking at payoffs signals selfish behavior, but it does not actually mean so. *Journal of Behavioral and Experimental Economics*, *65*, 79-84.

Critcher, C. R., Helzer, E. G., & Tannenbaum, D. (2020). Moral character evaluation: Testing another's moral-cognitive machinery. *Journal of Experimental Social Psychology, 87*, Article 103906.

Critcher, C.R., Inbar, Y., & Pizarro, D.A. (2013). How quick decisions illuminate moral character. *Social Psychological and Personality Science*, *4(*3*)*, 308-315.

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*(5), 979.

Ditto, P. H. & Liu, B. S. (2011). What Dilemma? Moral evaluation shapes factual belief. *Social Psychological and Personality Science, 4*(3), 316-323.

Effron, D. A., & Monin, B. (2010). Letting people off the hook: When do good deeds excuse transgressions?. *Personality and Social Psychology Bulletin*, 36(12), 1618-1634.

Effron, D. A., O'Connor, K., Leroy, H., & Lucas, B. J. (2018). From inconsistency to hypocrisy: When does "saying one thing but doing another" invite condemnation?. *Research in Organizational Behavior, 38*, 61-75.

Evans, A. M., & Van de Calseyde, P. P. (2017). The effects of observed decision time on expectations of extremity and cooperation. *Journal of Experimental Social Psychology*, *68*, 50 –59.

Everett, J. A., Pizarro, D. A., & Crockett, M. J. (2016). Inference of trustworthiness from intuitive moral judgments. *Journal of Experimental Psychology: General*, *145*(6), 772-787.

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

Fu, G., Xu, F., Cameron, C. A., Heyman, G., & Lee, K. (2007). Cross-cultural differences in children's choices, categorizations, and evaluations of truths and lies. *Developmental Psychology*, 43(2), 278-293.

Fu, G., Lee, K., Cameron, C. A., & Xu, F. (2001). Chinese and Canadian adults' categorization and evaluation of lie-and truth-telling about prosocial and antisocial behaviors. *Journal of Cross-Cultural Psychology*, 32(6), 720-727.

Galak, J., & Critcher, C. R. (2022). Who sees which political falsehoods as more acceptable and why: A new look at in-group loyalty and trustworthiness. *Journal of Personality and Social Psychology.*

Garrett, N., Lazzaro, S. C., Ariely, D., & Sharot, T. (2016). The brain adapts to dishonesty. *Nature neuroscience*, *19*(12), 1727-1732.

Gerlach, P., Teodorescu, K., & Hertwig, R. (2019). The truth about lies: A meta-analysis on dishonest behavior. *Psychological Bulletin*, 145(1), 1.

Gilbert, D. T., & Malone, P. S. (1995). The correspondence bias. *Psychological Bulletin*, 117(1), 21.

Gino, F. & Galinsky, A. D. (2012). Vicarious dishonesty: When psychological closeness creates distance from one's moral compass. *Organizational Behavior and Human Decision Processes, 119*(10, 15-26.

Gino, F., & Schweitzer, M. E. (2008). Blinded by anger or feeling the love: how emotions influence advice taking. *Journal of Applied Psychology*, 93(5), 1165.

Goffman, E. (1955). On face-work: An analysis of ritual elements in social interaction. *Psychiatry*, *18*(3), 213-231.

Goodwin, G. P., Piazza, J., & Rozin, P. (2014). Moral character predominates in person perception and evaluation. *Journal of Personality and Social Psychology, 106(*1), 148-168.

Graham, J., Meindl, P., Koleva, S., Iyer, R., & Johnson, K. M. (2015) When values and behavior conflict: Moral pluralism and intrapersonal moral hypocrisy. *Social and Personality Psychology Compass*, *9*(3), 158-170.

Gray, K., MacCormack, J. K., Henry, T., Banks, E., Schein, C., Armstrong-Carter, E., Abrams, S., & Muscatell, K. A. (2022). The affective harm account (AHA) of moral judgment: Reconciling cognition and affect, dyadic morality and disgust, harm and purity. *Journal of Personality and Social Psychology*, in press.

Gray, K., Schein, C., & Ward, A. F. (2014). The myth of harmless wrongs in moral cognition: Automatic dyadic completion from sin to suffering. *Journal of Experimental Psychology: General*, 143(4), 1600.

Greene, J. D. (2007). Why are VMPFC patients more utilitarian? A dual-process theory of moral judgment explains. *Trends in Cognitive Sciences, 11*(8), 322-323.

Greene, J. D., Sommerville, R. B., Nystrom, L. E., Darley, J. M., & Cohen, J. D. (2001). An fMRI investigation of emotional engagement in moral judgment. *Science, 293*(5537), 2105-2108.

Hart, J. L. (2022). Deception, honesty, and professionalism: a persistent challenge in modern medicine. *Current Opinion in Psychology*, 101434.

Hartley, A. G., Furr, R. M., Helzer, E. G., Jayawickreme, E., Velasquez, K. R., & Fleeson, W. (2016). Morality's centrality to liking, respecting, and understanding others. *Social Psychological and Personality Science, 7*(7), 648-657.

Haselton, M. G., & Buss, D. M. (2000). Error management theory: a new perspective on biases in cross-sex mind reading. *Journal of Personality and Social Psychology, 78*(1), 81.

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

Heyman, G. D., Hsu, A. S., Fu, G., & Lee, K. (2013). Instrumental lying by parents in the US and China. *International Journal of Psychology*, 48(6), 1176-1184.

Hildreth, J. A. D., & Anderson, C. (2018). Does loyalty trump honesty? Moral judgments of loyalty-driven deceit. *Journal of Experimental Social Psychology, 79*, 87-94.

Hildreth, J. A.D., Gino, F., & Bazerman, M. (2016). Blind loyalty? When group loyalty makes us see evil or engage in it. *Organizational Behavior and Human Decision Processes, 132*, 16-36.

Hoffman, M., Yoeli, E., & Nowak, M.A. (2015). Cooperate without looking: Why we care what people think and not just what they do. *Proceedings of the National Academy of Sciences,* 1–6. DOI:10.1073/pnas.1417904112.

Hofmann, W., Wisneski, D. C., Brandt, M. J., & Skitka, L. J. (2014). Morality in everyday life. *Science, 345*(6202), 1340-1343.

Holyoak, K. J., & Powell, D. (2016). Deontological coherence: A framework for commonsense moral reasoning. *Psychological Bulletin*, 142(11), 1179.

Huppert, E., Herzog, N., Levine, E.E., Landy, J. (2023). On being dishonest about dishonesty: The social costs of taking nuanced (but realistic) moral stances. *Journal of Personality and Social Psychology.* Advance online publication. https://doi.org/10.1037/pspa0000340

Jampol, L., & Zayas, V. (2021). Gendered White Lies: Women Are Given Inflated Performance Feedback Compared With Men. *Personality and Social Psychology Bulletin, 47*(1), 57-69.

Jordan, J.J., Hoffman, M., Nowak, M.A., & Rand, D.G. (2016). Uncalculating cooperation is used to signal trustworthiness. *Proceedings of the National Academy of Sciences*, *113*(31), 8658-8663.

Jordan, J., Sommers, R., Bloom, P., & Rand, D. G. (2017). Why do we hate Hypocrites? Evidence for a Theory of False Signaling. *Psychological Science, 28*(3), 356-368.

Kahane, G., Everett, J. A., Earp, B. D., Caviola, L., Faber, N. S., Crockett, M. J., & Savulescu, J. (2018). Beyond sacrificial harm: A two-dimensional model of utilitarian psychology. *Psychological Review*, 125(2), 131.

Kahane, G., Wiech, K., Shackel, N., Farias, M., Savulescu, J., & Tracey, I. (2012). The neural basis of intuitive and counterintuitive moral judgment. *Social Cognitive and Affective Neuroscience, 7*(4), 393-402.

Kahneman, D. & Tversky, A. (1979). Prospect Theory: An Analysis of Decision under Risk *Econometrica, 47*(2), 263-292.

Kidder, R. M. (1995). *How good people make tough choices*. New York: Morrow.

Kohlberg, L. 1971. Stages of moral development as a basis for moral education. In C. M. Beck, B. S. Crittenden & E. V. Sullivan (Eds.), Moral education: Interdisciplinary approaches: 23–92. Toronto, Canada: University of Toronto Press.

Kreps, T. A., & Monin, B. (2014). Core values versus common sense: Consequentialist views appear less rooted in morality. *Personality and Social Psychology Bulletin*, *40*(11), 1529-1542.

Krijnen, J. M., Tannenbaum, D., & Fox, C. R. (2017). Choice architecture 2.0: Behavioral policy as an implicit social interaction. *Behavioral Science & Policy*, *3*(2), i-18.

Kubin, E., von Sikorski, C., & Gray, K. (2022, June 8). When Censorship Feels Acceptable: People Suppress Political Ideas They Perceive as Harmful Lies. https://doi.org/10.31234/osf.io/ha8nv

Landy, J. F., & Uhlmann, E. L. (2018). Morality is personal. *Atlas of Moral Psychology*, 121.

Lee, J.J., Hardin, A. E., Parmar, B., & Gino, F. (2019). The interpersonal costs of dishonesty: How dishonest behavior reduces individuals' ability to read other' emotions. *Journal of Experimental Psychology: General*, *148*(9), 1557-1574.

Lee, K., Xu, F., Fu, G., Cameron, C. A., & Chen, S. (2001). Taiwan and Mainland Chinese and Canadian children's categorization and evaluation of lie-and truth-telling: A modesty effect. *British Journal of Developmental Psychology*, 19(4), 525-542.

Levine, E.E. (2022). Community standards of deception: Deception is perceived to be ethical when it prevents unnecessary harm. https://doi.org/10.31234/osf.io/g5trb

Levine, E. E., Barasch, A., Rand, D., Berman, J. Z., & Small, D. A. (2018). Signaling emotion and reason in cooperation. *Journal of Experimental Psychology: General*, *147*(5), 702-719.

Levine, E. E., & Cohen, T. R. (2018). You can handle the truth: Mispredicting the consequences of honest communication. *Journal of Experimental Psychology: General*, 147(9), 1400.

Levine, E. E., & Lupoli, M. J. (2021). Prosocial Lies: Causes and Consequences. *Current Opinion in Psychology, 43*, 335-340.

Levine, E.E. & Munguia Gomez, D. (2021). "I'm just being honest." When and why honesty enables helping versus harming behaviors. *Journal of Personality and Social Psychology, 120*(1): 33-56.

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.

Levine, E.E. & Schweitzer, M.E. (2015). Prosocial lies: when deception breeds trust. *Organizational Behavior and Human Decision Processes, 126*, 88-106.

Lipsey, M. W. & Wilson, D. B. (2001). *Practical meta-analysis*. Sage Publications, Inc.

Liu, B. S., & Ditto, P. H. (2013). What dilemma? Moral evaluation shapes factual belief. *Social Psychological and Personality Science*, *4*(3), 316-323.

Lupoli, M. J., Levine, E. E., & Greenberg, A. E. (2018). Paternalistic lies. *Organizational Behavior and Human Decision Processes*, *146*, 31-50.

Lupoli, M. J., Jampol, L., & Oveis, C. (2017). Lying because we care: Compassion increases prosocial lying. *Journal of Experimental Psychology: General*, 146(7), 1026.

Mazar, N., Amir, O., & Ariely, D. (2008). The Dishonesty of Honest People: A Theory of Self-Concept Maintenance. *Journal of Marketing Research, 45*(6), 633-644.

Nisbett, R. E., Peng, K., Choi, I., & Norenzayan, A. (2001). Culture and systems of thought: Holistic versus analytic cognition. *Psychological Review, 108*(2), 291–310.

Piazza, J. & Landy, J. (2013). "Lean not on your own understanding": Belief that morality is founded on divine authority and non-utilitarian moral judgments. *Judgment and Decision Making, 8*, 639-661.

Rand, D. G., Peysakhovich, A., Kraft-Todd, G. T., Newman, G. E., Wurzbacher, O., Nowak, M. A., & Greene, J. D. (2014). Social heuristics shape intuitive cooperation. *Nature Communications*, 5(1), 1-12.

Reed, A., & Aquino, K. F. (2003). Moral Identity and the Expanding Circle of Moral Regard towards Out-Groups. *Journal of Personality and Social Psychology*, 84, 1270-1286. http://dx.doi.org/10.1037/0022-3514.84.6.1270

Rogers, T., Zeckhauser, R., Gino, F., Norton, M. I., & Schweitzer, M. E. (2017). Artful paltering: The risks and rewards of using truthful statements to mislead others. *Journal of personality and social psychology*, *112*(3), 456.

Scheffler, S. (1982).The rejection of consequentialism. Oxford: Oxford University Press

Schein, C. & Gray, K. (2018). The Theory of Dyadic Morality: Reinventing Moral Judgment by Redefining Harm. *Personality and Social Psychology Review*, *22*(1), 32-70.

Sen, A. (1983). Evaluator relativity and consequential evaluation. *Philosophy and Public Affairs,* 12,113–132.

Shafir, E. & Tversky, A. (1992). The disjunction effect in choice under uncertainty. *Psychological Science, 3*(5), 305-310. DOI: 10.1111/j.1467-9280.1992.tb00678.x

Talwar, V., Murphy, S. M., & Lee, K. (2007). White lie-telling in children for politeness purposes. *International Journal of Behavioral Development, 31*(1), 1-11.

Tong, D., Isik, I., & Talwar, V. (2023). A cross-cultural comparison of the relation between children's moral standards of honesty and their lie-telling behavior. *Journal of Experimental Child Psychology*, 231, 105665.

Uhlmann, E. L., Zhu, L. (L.), & Tannenbaum, D. (2013). When it takes a bad person to do the right thing. *Cognition, 126*(2), 326–334.

Weisel, O., & Shalvi, S. (2015). The collaborative roots of corruption. *Proceedings of the National Academy of Sciences, 112*(34), 10651-10656.

Wiltermuth, S. S. (2011). Cheating more when the spoils are split. *Organizational Behavior and Human Decision Processes, 115*(2), 157-168.

Van de Calseyde, P. P., Keren, G., & Zeelenberg, M. (2014). Decision time as information in judgment and choice. *Organizational Behavior and Human Decision Processes, 125*(2), 113-122.

Van Zant, A. B. & Moore, D. A. (2015). Leaders' use of moral justifications increases policy support. *Psychological Science, 26* (6), 934-943.

Vrij, A. (2007). Deception: A social lubricant and a selfish act. Social communication, 309-342.

Zhang, T., Gino, F., & Margolis, J. D. (2018). Does "could" lead to good? On the road to moral insight. *Academy of Management Journal*, *61*(3), 857-895.

Zlatev, J. J. (2019). I may not agree with you, but I trust you: Caring about social issues signals integrity. *Psychological Science*, 30(6), 880-892.

Zlatev, J. J., Kupor, D. M., Laurin, K., & Miller, D. T. (2020). Being "good" or "good enough": Prosocial risk and the structure of moral self-regard. *Journal of Personality and Social Psychology*, 118(2), 242.