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

Supplemental Online Materials

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] David 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

Table of Contents

1. Additional results and information from reported studies

In the main manuscript, we report the results of our focal dependent variable: moral judgment and weight of advice (WOA; Study 2). Below, we provide supplementary details of our materials and methods, and we report all ancillary measures collected in our main studies. Unless otherwise noted, standard deviations appear in parentheses next to the means in tables with demographic information.

1.1. Study 1

1.1.1. Scenario text from Study 1

Table S1.

Scenario Text from Study 1

Known Conditions		Unknown Condition
Imagine an employee who must deliver an important presentation. He will pitch a new marketing plan to his manager and colleagues. He plans on wearing his favorite black suit during the presentation. Imagine that the employee's colleague – unbeknownst to the employee - thinks this suit is too tight and that the suit is inappropriate for the presentation. The day of his presentation, the employee shows up in his suit and he asks his colleague how he looks in it. **At this time, the employee has no other suits available that he can wear.** *(No time to change [Known: Honesty causes harm])*	Imagine an employee who must deliver an important presentation. He will pitch a new marketing plan to his manager and colleagues. He plans on wearing his favorite black suit during the presentation. Imagine that the employee's colleague – unbeknownst to the employee - thinks this suit is too tight and that the suit is inappropriate for the presentation. The day of his presentation, the employee shows up in his suit and he asks his colleague how he looks in it. **At this time, the employee has other suits available that he can wear.** *(Time to change [Known: Honesty helps])*	Imagine an employee who must deliver an important presentation. He will pitch a new marketing plan to his manager and colleagues. He plans on wearing his favorite black suit during the presentation. Imagine that the employee's colleague – unbeknownst to the employee - thinks this suit is too tight and that the suit is inappropriate for the presentation. The day of his presentation, the employee shows up in his suit and he asks his colleague how he looks in it. *(Unknown if honesty causes harm)*

1.1.2. Post-test of Study 1

1.1.2.1. Method for Post-test of Study 1

Participants
- *N* = 100 (M_{age} = 39.62, SD_{age}= 12.18; 57 male, 42 female, 1 prefer to self-describe)
- Collected on MTurk

Procedure
- This study featured the same scenarios as Study 1 (see Table S1 above). Participants were all assigned to the same (Looking) condition.
- Participants read about a colleague and employee. The employee wears a suit to a big presentation, and the colleague thinks the suit is inappropriate. When the employee asks the colleague what he thinks of the suit, all participants read that the colleague "dodges the question" and looks for more information. We used this statement to operationalize "Looking" (i.e., conditional honesty).

Dependent Variables

- **Morality**. As in our main studies, participants made judgments of the Communicator's morality using the same three item composite as in Studies 2-5 (α = .931).
- **General Lying.** We asked participants to indicate their beliefs about two items: "In general, how often do you think [the Communicator] tells lies that help others?" and In general, how often do you think [the Communicator] tells lies that harm others?" (1 = Never, 7 = Very often).
- **Likelihood of truth-telling and lying if target could implement feedback.** We asked participants to indicate what the communicator would do if the target had time to implement feedback. Specifically, we asked "If the colleague learned that the employee could change his suit, how likely do you think it is that: 1) the colleague would tell the truth: saying that he thinks the suit is inappropriate, or 2) the colleague would lie: saying that he thinks the suit is fine." (Participants responded to both items using a continuous slider from 0-100).
- **Leakiness of Looking.** We asked participants the extent to which they thought the target was able to figure out the communicator's true opinion of his suit based on the communicator's looking behavior (1=Not at all, 7=A great extent).
- **Attributions of harmful lies.** We asked participants to imagine that the communicator told a harmful lie, and then rate the extent to which ten different attributions could explain that behavior. Specifically, the attributions were that 1) the [communicator] wants to undermine the employee, 2) the [communicator] wants to be polite, 3) the [communicator] is lazy, 4) the [communicator] is insensitive, 5) the [communicator] doesn't care about others, 6) the [communicator] doesn't have enough self-control to tell the truth, 7) the [communicator] is a mean person, 8) the [communicator] is a dishonest person, 9) the [communicator] is an unprincipled person, and 10) the [communicator] wants to avoid conflict.

1.1.2.2. Results and Discussion for Post-test of Study 1

Results

Table S2 reports the descriptive statistics of all measures. In addition to examining the descriptive statistics, we ran a regression entering the variables 'tells-lies-to-help-others' and 'tells-lies-to-harm-others' as the independent variables and morality as the dependent variable. The belief that the communicator would tell lies to help others was significantly and positively related to morality (*b* = 0.22,

SE = 0.09, *p* = .012). The belief that the communicator would tell lies to harm others, however, was negatively related to morality (*b* = -0.22, SE = 0.07, *p* = .004).

Table S2.

Descriptive statistics for the dependent measures reported in the post-test of Study 1

	M	*SD*
Morality	5.10	1.22
General Lying		
In general, how often do you think [the Communicator] tells lies that harm others?	3.10	1.63
In general, how often do you think [the Communicator] tells lies that help others?	4.29	1.40
Likelihood[1]		
Tell-truth-if-could-change	71.13	24.34
Tell-lie-if-could-change	33.65	29.31
Leakiness of Looking		
To what extent do you think the employee was able to figure out that the colleague had a negative opinion of his suit based on these actions?	4.34	1.36
Attributions		
The [Communicator] wants to undermine the employee	3.27	1.82
The [Communicator] wants to be polite	5.21	1.67
The [Communicator] is lazy	2.74	1.69
The [Communicator] is insensitive	3.35	1.95
The [Communicator] doesn't care about others	3.49	1.87
The [Communicator] doesn't have enough self-control to tell the truth	3.52	1.76
The [Communicator] is a mean person	3.12	1.77
The [Communicator] is a dishonest person	3.85	1.78
The [Communicator] is an unprincipled person	3.51	1.86
The [Communicator] wants to avoid conflict	5.32	1.60

---

[1] These items were not presented as mutually exclusive in the survey, even though they should technically add to 100%.

1.2. Study 2

1.2.1. Additional measures and results in Study 2

In addition to the measures reported in the manuscript, participants reported their judgments of the Communicator's trustworthiness (composite of "I trust [Communicator]," "I am willing to make myself vulnerable to [Communicator]," and "I am confident that [Communicator] sent me honest advice"; $\alpha = .914$).

We ran an independent samples t-test on perceived trustworthiness using Decision Strategy (i.e., Unconditional Honesty vs. Looking) as the independent variable. Participants perceived the Communicator who was unconditionally honest ($M = 3.68$, $SD = 1.59$) as somewhat, but not significantly, more trustworthy than the Communicator who looked at the consequences ($M = 3.49$, $SD = 1.54$; $t(238) = .975$, $p = .331$).

1.3. Study 3

1.3.1. Additional measures and results in Study 3

We also explored potential attributions for telling a harmful lie within the Coin Flip Game. These questions were only asked in the Looking condition. Following the main dependent variables reported in the manuscript, participants in the Looking condition were asked to imagine that the Communicator chose to lie and reported that the coin flip landed on TAILS, causing another person to lose $1. Then, we measured 9 potential attributions for telling the harmful lie on a 1 (strongly disagree) to 7 (strongly agree) scale: "To what extent do you think each of the following explanations below explains the [Communicator]'s choice in this case?: "The [Communicator] did not realize that reporting TAILS would cause the RECEIVER to lose $1", "The [Communicator] did not have enough self-control to tell the truth", "The [Communicator] wanted to undermine the RECEIVER", "The [Communicator] was being lazy", "The [Communicator] was being insensitive", "The [Communicator] was not thinking much about the situation", "The [Communicator] is a mean person", "The [Communicator] is dishonest", and "The [Communicator] is unprincipled".

At the conclusion of the study, we also measured participants' general propensity to trust others (adapted from Ben-Ner & Halldorsson, 2010) using four items on a 1 (strongly disagree) to 7 (strongly agree) scale: "You cannot be too careful when dealing with people (*reverse scored*)", "Most people can be trusted", "I will not trust a person until there is clear evidence that he or she can be trusted (*reverse scored*)", and "I will trust a person until I have clear evidence that he or she can't be trusted".

As preregistered for exploratory purposes, we ran descriptive statistics on the attributions for telling a harmful lie. We report these descriptive statistics below in Table S3.

Table S3.

Descriptive statistics of potential attributions for telling a harmful lie within the Coin Flip Game (asked in Looking condition only)

Attribution	M	SD
The [Communicator] did not realize that reporting TAILS would cause the RECEIVER to lose $1	2.88	1.54
The [Communicator] did not have enough self-control to tell the truth	3.76	1.70
The [Communicator] wanted to undermine the RECEIVER	5.14	1.57
The [Communicator] was being lazy	3.21	1.55
The [Communicator] was being insensitive	5.13	1.48
The [Communicator] was not thinking much about the situation	3.87	1.67
The [Communicator] is a mean person	5.00	1.53
The [Communicator] is dishonest	5.55	1.43
The [Communicator] is unprincipled	5.15	1.38

As preregistered for exploratory purposes, we also explored whether the effect of Decision Strategy on judgments of morality, and beliefs about harmful and prosocial lies, are moderated by participants' general propensity to trust others. Specifically, we conducted a linear regression on judgments of morality, and the four lying judgments (tendency to tell harmful lies, tendency to tell helpful lies, probability of telling a harmful lie in the Coin Flip Game, and probability of telling a helpful lie in

the Coin Flip Game) using Decision Strategy (1 = Unconditional Honesty, -1 = Looking), General Propensity to Trust, and their interactions as independent variables.

We did not detect a significant interaction of Decision Strategy and the General Propensity to Trust on judgments of morality (p = .810). Thus, we did not find evidence that the effect of Unconditional Honesty (vs. Looking) on judgments of morality is moderated by people's general propensity to trust others.

We also did not detect significant interactions of Decision Strategy and the General Propensity to Trust on the tendency to tell helpful lies in general (p = .502), the probability of telling a harmful lie in the Coin Flip Game (p =.374), or the probability of telling a helpful lie in the Coin Flip Game (p = .899).

We did, however, find a significant interaction of Decision Strategy and the General Propensity to Trust on the belief that the Communicator tells harmful lies in general (b = 0.08, SE = 0.04, p = .047). When Communicators engaged in Unconditional Honesty, participants' General Propensity to Trust did not predict their belief that the Communicator has the tendency to tell harmful lies (p =.678). However, when Communicators engaged in Looking, participants' General Propensity to Trust was negatively associated with their belief that the Communicator has the tendency to tell harmful lies (b = -0.19, SE = 0.06, p =.003). This moderation reveals that the propensity to trust others may buffer against the concern that Communicators who look for more information will tell harmful lies in general.

1.4. Study 4

1.4.1. Additional measures and results in Study 4

At the conclusion of the study, we asked participants to indicate their opinions about why rules exist using six items adapted from Piazza & Landy (2013): "Rules exist because people would often make mistakes or bad decisions without them," "Rules exist because human judgment is flawed," "Rules exist because people do not always know what the best course of action is," "Rules exist to prevent people from acting upon their nature, sinful impulses," "Rules exist because people's depraved nature prevents them from making moral decisions on their own," and "Rules exist because people will naturally choose wickedness over goodness (1= Strongly disagree, 9=Strongly agree; α = .826).

These measures capture the degree to which people believe, in general, that rules are useful for preventing unethical actions from occurring. We measured them with the intention of exploring whether they moderate the effect of Unconditional Honesty on judgments of morality. Therefore, we conducted a linear regression on judgments of morality using Decision Strategy (1 = Unconditional Honesty, -1 = Looking), Human Fallibility, and their interaction as independent variables. We do not find a significant interaction between Decision Strategy and Human Fallibility ($p = .618$).

1.5. Study 5

1.5.1. Additional measures and results in Study 5

For exploratory purposes, we also preregistered the measurement of two alternative mechanisms. Specifically, we examined the extent to which participants believed that Communicators who Looked would go down a "slippery slope" of deceptive behavior in the future (Anderson et al., 2023; Garrett, Lazzaro, Ariely, & Sharot, 2016), and that they had poor internalized honesty standards (Huppert et al., 2023). We measured slippery slope beliefs (using two items: "I'm concerned that [the Communicator's] current behavior may lead them to tell bigger lies in the future" and "I'm concerned that [the Communicator's] current behavior may make it easier for them to justify future lies" ($r(770) = .794$). We examined judgments of the Communicator's internalized honesty standards, independent of (future) lie behavior using three-items: "The [Communicator] would feel guilty if they lied," "The [Communicator] thinks carefully before lying," and "The [Communicator] genuinely values honesty" ($\alpha = .753$).

We conducted a two-way ANOVA on beliefs that the Communicator's behavior would lead to a slippery slope of deceptive behavior in the future using Decision Strategy and Decision Option as factors. This analysis revealed a main effect of Decision Strategy ($F(1,766) = 59.72$, $p < .001$). Communicators who engaged in Unconditional Honesty were judged as less likely to be on a slippery slope of deceptive behavior compared to those who Looked. We did not detect a main effect of Decision Option on these beliefs ($p = .107$), nor a significant interaction between Decision Strategy and Decision Option ($p = .697$). We report the corresponding descriptive statistics in Table S5.

We also conducted a two-way ANOVA on judgments of the Communicator's internalized honesty standards using Decision Strategy and Decision Option as factors. This revealed a main effect of Decision Strategy ($F(1,766) = 96.84$, $p < .001$) such that Unconditionally Honest Communicators were judged as having greater internalized honesty standards than those who Looked. There was no significant effect of Decision Option on this judgment ($p = .388$), nor a significant interaction between Decision Strategy and Decision Option ($p = .605$). The corresponding descriptive statistics are reported in Table S5.

Table S5.

Descriptive statistics for slippery slope of deceptive behavior and internalized honesty standards

	No Precommitment		Precommitment	
Variable	**Unconditional Honesty**	**Looking**	**Unconditional Honesty**	**Looking**
Slippery Slope of Deceptive Behavior	2.16 (1.34)	3.00 (1.43)	2.04 (1.17)	2.79 (1.67)
Internalized Honesty Standards	5.17 (1.08)	4.42 (1.02)	5.14 (1.13)	4.31 (1.19)

As preregistered, we tested 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, tendency to tell prosocial lies, beliefs about a slippery slope of deceptive behavior, and internalized honesty standards as simultaneous mediators. We report moderated mediation results with only our focal mechanisms (tendency to tell harmful vs. prosocial lies) in the main manuscript, per the guidance of the review team.

Results of the moderated mediation analysis with all four mechanisms are presented in Table S6. Consistent with our theorizing, we found significant moderated mediation through beliefs about the

Communicator's tendency to tell harmful lies (index of moderated mediation = -0.35, SE = 0.07, 95% CI [-0.50, -0.23]). Engaging in Unconditional Honesty led to lower concerns about harmful lies, which led to higher judgments of morality (indirect effect = 0.24, SE = 0.05, 95% CI [0.16, 0.34]), when Communicators who Looked did not express precommitment. In contrast, when Communicators who Looked precommitted to making a prosocial decision, harmful lies mediated the effect of Unconditional Honesty on judgments of morality in the opposite direction; engaging in Unconditional Honesty led to higher concerns about harmful lies, which led to lower ratings of morality (indirect effect = -0.11, SE = 0.04, 95% CI [-0.20, -0.03]).

We also found significant moderated mediation through beliefs about the Communicator's tendency to tell prosocial lies (index of moderated mediation = -0.10, SE = 0.03, 95% CI [-0.18, -0.05]). When Communicators who Looked did not precommit, prosocial lies mediated the effect of Unconditional Honesty on judgments of morality; Unconditional Honesty led to lower beliefs about the propensity to tell prosocial lies, which led to lower ratings of morality (indirect effect = -0.14, SE = 0.03, 95% CI [-0.20, -0.08]). However, the indirect effect of prosocial lies was significantly larger when Communicators who Looked precommitted to making a prosocial decision (indirect effect = -0.24, SE = 0.05, 95% CI [-0.35, -0.14]).

Although we also found evidence of mediation through internalized honesty standards (consistent with Huppert et al., 2023), the strength of mediation did not depend on the Communicator's precommitment. These results suggest that concerns about overall honesty standards may be independent of concerns about the risks of harmful lies. We find no evidence of mediation through slippery slope beliefs.

Table S6.

Supplemental Moderated Mediation Analysis from Study 5

Mediators	No Precommitment	Precommitment	Index of Moderated Mediation
Tendency to tell harmful lies	**0.16, 0.34**	**-0.20, -0.03**	**-0.50, -0.22**
Tendency to tell prosocial lies	**-0.20, -0.08**	**-0.35, -0.14**	**-0.18, -0.05**
Slippery slope beliefs	0.00, 0.10	0.00, 0.09	-0.04, 0.02
Internalized honesty standards	**0.25, 0.49**	**0.28, 0.55**	-0.11, 0.19

2. **Supplemental Studies**

Across our supplemental studies, participants gave electronic consent and had to pass a series of attention checks before beginning the survey. Those who did not successfully complete the checks were informed that they could not continue with the study, and their data was not included in the final analysis. Unless otherwise noted, standard deviations appear in parentheses next to the means in tables with demographic information.

2.1. Study S1: Moral Judgments of Unconditional Honesty Within Subjects

In Study S1, we conceptually replicate Study 1 using a within-subjects, rather than between-subjects, design. In doing so, Study S1 ensures that the preference for unconditional honesty is not explained by a failure to think through the decision tree (Shafir & Tversky, 1992). When the social consequences of honesty are unknown, people may simply fail to consider what is ethical if honesty were helpful versus harmful. We rule out this possibility in Study S1; participants preferred Unconditional honesty (to Looking) when consequences were unknown even *after* they made moral judgments about honesty that would help or harm another person.

2.1.1. Method for Study S1

Participants
- $N = 168$, 167 of which provided at least some demographic information ($M_{age} = 33.30$, $SD_{age} = 15.82$; 74 males, 93 females, 1 did not report)
- Collected from local city parks[2]

Procedure
- Participants read a scenario in which a Communicator faced the decision to reveal that they had cheated on a romantic partner (see Table S7 for exact text).
- Participants were assigned to one of two conditions in a between-subjects design (Condition: Unknown vs. Known-Followed-By-Unknown).
- The Unknown condition was identical in design to the Unknown condition in Study 1. Participants made a single judgment about whether to engage in honesty or lie when the consequences of honesty and lying were not known.
- In the Known-Followed-By-Unknown condition, participants made three judgments about whether to engage in honesty. They made two judgments when the consequences were known, identical in design to the Known condition in Study 1, but after this, they then made a third judgment when the consequences were unknown.

Dependent Variables
- Our main dependent variable was the judgment about whether to engage in honesty, lie, or look.
- We also measured the perceived ethicality of each of the following actions (1 = Very unethical, 7 = Very ethical): "Lying to Jo about being unfaithful when Jo's sickness is treatable (Jo will live [Harmful lie])," "Telling the truth to Jo about being unfaithful when Jo's sickness is treatable (Jo will live [Prosocial truth])," Lying to Jo about being unfaithful when Jo's sickness is untreatable (Jo will die [Prosocial lie])," and "Telling the truth to Jo about being unfaithful when Jo's sickness is untreatable (Jo will die [Harmful truth])."
- Participants also rated their agreement with the following statements (1 = Strongly disagree, 7 = Strongly agree): "Telling the truth without finding out more information ensures that Sam does not make a really bad decision," and "A person who would tell the truth without finding out more information is probably more honest across situations."

---

[2] We also recruited participants from our behavioral laboratory. The researchers have the raw data but did not have access to the source file at the time of publication. For that reason, we report only the field data here. If the researchers gain access to the source file for the laboratory data in the future, these results will be updated.

Table S7.

Scenario Text from Study S1

Consequences-Known-Followed-By-Consequences-Unknown Condition	Consequences-Unknown-Only Condition
Imagine the following scenario:	Imagine the following scenario:
Jo and Sam are a married couple. A few months ago, Sam cheated on Jo. It only happened once, and Sam remains very much in love with Jo. Jo does not know that Sam cheated.	Jo and Sam are a married couple. A few months ago, Sam cheated on Jo. It only happened once, and Sam remains very much in love with Jo. Jo does not know that Sam cheated.
A few weeks after Sam cheated, Jo was diagnosed with a very serious, life-threatening illness.	A few weeks after Sam cheated, Jo was diagnosed with a very serious, life-threatening illness.
When Jo was diagnosed, Jo's doctor said there is a 50% chance that the sickness is treatable and Jo will live and a 50% chance that the sickness is untreatable and Jo will die in 1 week.	When Jo was diagnosed, Jo's doctor said there is a 50% chance that the sickness is treatable and Jo will live and a 50% chance that the sickness is untreatable and Jo will die in 1 week.
Jo's doctor just completed a test and now knows whether Jo's sickness is treatable.	Jo's doctor just completed a test and now knows whether Jo's sickness is treatable.
Jo knows the test results, but <u>Sam does not</u>. *(Consequences-Known Untreatable [Honesty causes harm])* Now imagine that Jo and Sam both know the test results. <u>Jo's sickness is not treatable. Jo is likely to die in 1 week.</u> *(Consequences-Known Treatable [Honesty helps])* Now imagine that Jo and Sam both know the test results. <u>Jo's sickness is treatable. Jo is likely to live.</u> *(Consequences-Unknown)* Now imagine that Jo knows the test results, but <u>Sam does not.</u>	**Jo knows the test results, but <u>Sam does not</u>.** *(Consequences-Unknown-Only)*

2.1.2. Results & Discussion for Study S1

Ethical Choice

In the Unknown condition, 76.3% of participants chose Unconditional Honesty as the most ethical choice, 10.0% chose Looking, and 13.8% chose Unconditional Lying. Consistent with Study 1, the majority of participants saw Unconditional Honesty as the most ethical choice when the consequences of honesty were unknown. This distribution was similar to the distribution of choices in the Unknown phase of the Known-Followed-by-Unknown condition (69.9% Unconditional Honesty, 15.7% Looking, 14.5%

Unconditional Lying; $\chi^2 = 1.26$, $p = .534$). Therefore, across both conditions, participants believed unconditional honesty was most ethical when the consequences of lying and truth-telling were unknown.

Notably, however, many of the same participants who indicated that Unconditional Honesty was the most moral choice when the consequences of honesty were unknown, also indicated that deception was acceptable when the consequences of honesty were known. That is, the distribution of choices in the Unknown phase of the Known-Followed-by-Unknown condition significantly differed from the distribution of choices expressed by these *same* participants within the Known portion of the Known-Followed-by-Unknown condition (69.9% preferred Unconditional Honesty, 15.7% preferred Looking, and 14.5% preferred Unconditional Lying; $z = -4.53$, $p < .001$).[3]

Other Measures

To explore which of these potential mechanisms are related to people's preference for Unconditional Honesty, we conducted a linear regression on people's voiced decision under uncertain consequences (1 = Unconditional Honesty, 0 = Other (i.e., Lie or Looking)) using the six items above as independent variables. Table S8 displays the descriptive results for each item, and the regression results appear in Table S9.

Table S8.

Descriptive Statistics of Exploratory Mechanism Items in Study S1

	M	*SD*
Ethicality of: Lying to Jo about being unfaithful when Jo's sickness is treatable (Jo will live [Harmful lie])	1.81	1.25
Ethicality of: Telling the truth to Jo about being unfaithful when Jo's sickness is treatable (Jo will live [Prosocial truth])	4.88	2.04
Ethicality of: Lying to Jo about being unfaithful when Jo's sickness is untreatable (Jo will die [Prosocial lie])	2.94	1.78
Ethicality of: Telling the truth to Jo about being unfaithful when Jo's sickness is untreatable (Jo will die [Harmful truth])	3.94	1.83
Agreement with: Telling the truth without finding out more information ensures that Sam does not make a really bad decision	4.49	1.79
Agreement with: A person who would tell the truth without finding out more information is probably more honest across situations	5.74	1.47

Table S9.

Regression of Exploratory Mechanisms Predicting People's Implied Preference when Consequences were Unknown in Study S1

---

[3] We note that our preregistration stated we would use a chi-square test to examine the (implied) distribution of choices within the "Unknown" and the "Known portions of the Known-Followed- by-Unknown condition. However, given that this was a within-subjects comparison, a chi-square test was not appropriate. We conducted a Wilcoxon Signed Ranks test instead.

	b	SE	p
(Constant)	0.64	0.17	<.001
Ethicality of: Lying to Jo about being unfaithful when Jo's sickness is treatable (Jo will live [Harmful lie])	**-0.06**	**0.03**	**.055**
Ethicality of: Telling the truth to Jo about being unfaithful when Jo's sickness is treatable (Jo will live Prosocial truth])	-0.01	0.02	.784
Ethicality of: Lying to Jo about being unfaithful when Jo's sickness is untreatable (Jo will die [Prosocial lie])	**-0.09**	**0.02**	**< .001**
Ethicality of: Telling the truth to Jo about being unfaithful when Jo's sickness is untreatable (Jo will die [Harmful truth])	0.03	0.02	.282
Agreement with: Telling the truth without finding out more information ensures that Sam does not make a really bad decision	**0.05**	**0.02**	**.015**
Agreement with: A person who would tell the truth without finding out more information is probably more honest across situations	0.03	0.03	.243

Notes. Regression model predicts people's voiced decision under uncertain consequences (Unconditional Honesty= 1; Other = 0), using above items as simultaneous predictors.

Discussion

　　Study S1 suggests that the preference for unconditional honesty is not driven by a failure to think through the decision tree and consider what is most ethical when honesty causes harm. People believe that unconditional honesty is ethical when the social consequences of honesty are unknown, even after they made judgments about what was 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.

　　These results also reveal that the degree to which participants believed it was ethical to tell the prosocial lie and harmful lie both negatively predicted the preference for unconditional honesty, though harmful lying did not reach statistical significance. The belief that unconditional honesty would prevent a bad decision ("Telling the truth without finding out more information ensures that Sam does not make a really bad decision") also predicted the preference for unconditional honesty, suggesting that the preference for unconditional honesty is driven in part by beliefs that the strategy prevents the worst moral outcomes, consistent with our theorizing.

2.2. Study S2: Judgments of Unconditional Honesty when Honesty has Personal vs. Social Consequences

In Study S2, we extend our investigation by examining whether people's judgments of communicators who engage in Unconditional Honesty versus Looking are influenced by whether honesty has consequences for the communicator (personal consequences) or the recipient of the lie (social consequences). In doing so, we manipulate whether the communicator faces a right-wrong dilemma or a right-right dilemma. When honesty only affects the communicator, the only temptation to "look" involves self-interest. The communicator might be tempted to lie to help themselves, thereby facing a right-wrong dilemma. On the other hand, when honesty affects the recipient of the lie, the temptation to "look" involves prosociality. The communicator might be tempted to lie to help another person, thereby facing a right-right dilemma.

2.2.1. Method for Study S2

Participants
- $N = 796$ ($M_{age} = 37.10$, $SD_{age} = 11.69$; 383 males, 412 females, 1 prefer to self-describe)
- Collected on MTurk

Procedure
- Participants judged and interacted with a Communicator who had previously participated in the Coin Flip Game (see Table S10) with a different participant.[4]
- We randomly assigned participants to a condition from a 2 (Decision Strategy: Unconditional Honesty vs. Looking) x 2 (Consequences: Personal vs. Social) between-subjects design.
- As in Study 2, participants learned that their assigned Communicator started the game without knowing the payouts associated with each outcome of the coin flip (i.e., the Communicator did not know the specific value associated with an outcome of heads ($A) or tails ($B).
- We also manipulated whether the Communicator's decision in the Coin Flip Game produced a payout for the Communicator or a Target (who was not the focal participant). In this iteration of the game, neither our focal participants, nor the Communicators, learned the possible values of $A and $B.
- Participants learned that the Communicator learned that the coin landed on heads and then either honestly reported HEADS without looking for information on the values of A and B (Unconditional Honesty condition) or looked for this information before reporting the outcome (Looking condition).

---

[4]In Study S2, participants observed and were matched with the decisions of real communicators (see SOM Study S3a for details).

MORAL JUDGMENTS OF UNCONDITIONAL HONESTY 19

Table S10.

Experimental Stimuli from Study S2 (Coin Flip Game Instructions)

Consequences for Other	Consequences for Self
Below, are the exact instructions that your partner saw as a **DECIDER.**	Below, are the exact instructions that your partner saw as a **DECIDER.**
COIN FLIP GAME - DECIDER INSTRUCTIONS In this task, you are in the role of **DECIDER**. You will be matched with a partner in the role of **RECEIVER**.	**COIN FLIP GAME - DECIDER INSTRUCTIONS** In this task, you are in the role of **DECIDER**.
As a **DECIDER**, you will be responsible for making a decision that determines the **RECEIVER's** bonus payment. Your decision will not affect your own bonus payment.	As a **DECIDER**, you will be responsible for making a decision that determines your own bonus payment. Your decision will not affect anyone else.
Specifically, you will have to report the outcome of a digital coin flip. Like a regular coin, the digital coin has a 50% chance of landing on heads and a 50% chance of landing on tails.	Specifically, you will have to report the outcome of a digital coin flip. Like a regular coin, the digital coin has a 50% chance of landing on heads and a 50% chance of landing on tails.
If you report that the coin landed on **HEADS**, your partner will receive **$A**	If you report that the coin landed on **HEADS**, you will receive **$A**
If you report that the coin landed on **TAILS**, your partner will receive **$B**	If you report that the coin landed on **TAILS**, you will receive **$B**
You do not know the numbers associated with **$A** and **$B**.	You do not know the numbers associated with **$A** and **$B**.
In this game, you will begin by flipping the coin. Then, you will learn the outcome of the coin flip.	In this game, you will begin by flipping the coin. Then, you will learn the outcome of the coin flip.
Then, you will have to select one of the following options • The coin landed on **HEADS** • The coin landed on **TAILS** • I'd like to find out the values of **$A** and **$B** before making my decision. (Once you know the numbers associated with $A and $B, you will then have to report that coin either landed on **HEADS** or **TAILS**)	Then, you will have to select one of the following options • The coin landed on **HEADS** • The coin landed on **TAILS** • I'd like to find out the values of **$A** and **$B** before making my decision. (Once you know the numbers associated with $A and $B, you will then have to report that coin either landed on **HEADS** or **TAILS**)
Press next to flip the coin.	Press next to flip the coin.

Dependent Variables
- Our main dependent variable was judgment of the Communicator's morality ($\alpha = .972$) using the same items as prior studies.
- Trust behavior was also measured using the pass rate in the Trust Game (1 = pass, 0 = keep), which we adapted from Levine & Schweitzer (2015). In this version of the Trust Game, our participants started the game with \$2 and chose whether they would like to "Keep \$2" for themselves or "Pass \$2" to the Communicator they just learned about. If the participant chose to "Keep \$2," the participant would earn \$2, and the Communicator would earn \$0. If the participant chose to "Pass \$2," the money would increase to \$5, and the Communicator could then choose to either "Keep \$5" or "Return \$2.50" to the participant (see OSF for full experimental stimuli). The decision to "Pass \$2" in the Trust Game was our primary measure of trusting behavior.
- We also measured judgments of the Communicator's trustworthiness (3-items; "I trust [the Communicator] to return the money," "I am confident [the Communicator] will return the money," and "[The Communicator] is trustworthy"; $\alpha = .938$), benevolence (2-items; "kind" and "benevolent"; $r = .788$), integrity ("consistent," "committed," and "principled"; $\alpha = .894$), and honesty (1-item; "honest").
-

Analyses
- We ran a 2x2 ANOVA on judgments of morality, trustworthiness, benevolence, integrity, and honesty using Decision Strategy and Consequences as factors.
- We also ran a logistic regression on passing behavior in the Trust Game (1 = Pass \$2, 0 = Keep \$2) as the DV using Decision Strategy (1 = Unconditional Honesty, 0 = Looking), Consequences (1 = Personal, 0 = Social), and their interactions as IVs.

2.2.2. Results & Discussion for Study S2

Morality

We found a main effect of Decision Strategy ($F(1,792) = 157.03$, $p < .001$) on perceived morality; Communicators who were unconditionally honest were viewed as more moral than Communicators who engaged in looking. We found no main effect of Consequences ($p = .579$). Surprisingly, we also did not find a Decision Strategy x Consequences interaction ($p = .246$). In other words, Unconditional Honesty was seen as more moral than Looking, regardless of whether the consequences of honesty affected the Communicator or others. The descriptive results are presented in Table S11.

Other Measures

We found a main effect of Decision Strategy on perceived trustworthiness ($F(1,792) = 46.06$, $p <$ .001), integrity ($F(1,792) = 130.88$, $p < .001$), benevolence ($F(1,792) = 70.77$, $p < .001$), and honesty ($F(1,792) = 245.19$, $p < .001$) such that Communicators who were honest without looking were rated more positively on these judgments than Communicators who looked at the consequences of honesty before making a decision.

We also found a main effect of Consequences on trustworthiness ($F(1,792) = 5.46$, $p = .020$) such that Communicators who made a decision that impacted others were viewed as more trustworthy than Communicators who made a decision that impacted themselves. We found no main effects of Consequences on integrity, benevolence, or honesty ($ps \geq .397$). We also found no significant Decision Strategy x Consequences interaction on trustworthiness, integrity, benevolence, or honesty ($ps \geq .142$). Table S11 depicts all descriptive statistics for our attitudinal measures.

The effect of Decision Strategy in the logistic regression on trust behavior in the Trust Game was insignificant at the $p < .05$ threshold ($p = .064$), though directionally it is worth noting that participants were somewhat more trusting of Communicators in the Unconditional Honesty (vs. Looking) condition. There was no significant effect of Consequences nor significant interaction of Decision Strategy and Consequences ($ps > .446$). Regression results are presented in Table S12.

Table S11.

Descriptive Statistics of Attitudinal Measures in Study S2

Dependent Variables	Personal Consequences		Social Consequences	
	Unconditional Honesty	Looking	Unconditional Honesty	Looking
Morality	5.33 (1.11)	4.21 (1.25)	5.28 (1.13)	4.35 (1.11)
Trustworthiness	4.00 (1.71)	3.03 (1.53)	4.12 (1.83)	3.47 (1.68)
Integrity	5.21 (1.03)	4.30 (1.18)	5.12 (1.06)	4.27 (1.01)
Benevolence	4.87 (1.24)	4.02 (1.27)	4.81 (1.12)	4.21 (1.19)
Honesty	5.56 (1.20)	4.00 (1.48)	5.45 (1.22)	4.14 (1.23)
Trust Behavior	53.8%	40.5%	53.8%	44.2%

Note. The descriptives reported here come from a MANOVA with all dependent variables, and therefore reflect descriptives only from the set of participants who responded to ALL variables. In the meta-analysis reported in the main manuscript, we report descriptives of morality, based on all participants who responded to that scale only, which leads to slight changes in overall descriptives.

Table S12.

Regression Results of Trust Behavior in Study S2

	b	SE	p
Constant	-0.233	0.15	.111
Decision Strategy	0.385	0.21	.064
Consequences	-0.153	0.20	.446
Decision Strategy x Consequences	0.155	0.286	.588

Notes. Regression model predicts trust behavior (i.e., Pass \$2 = 1, Keep \$2 = 0) using Decision Strategy (Unconditional Honesty = 1, Looking = 0) and Consequences (Personal = 1, Social = 0) and their interaction as predictors.

Discussion

Study S2 provides further support that people make more favorable ethical judgments of those who are unconditionally honest. Interestingly, we also find that these moral evaluations do not depend on whether the consequences will affect others or oneself. These results suggest an ethical preference for unconditional honesty regardless of who is affected by the potential consequences of the truth.

2.3. Study S3: Judgments of Unconditional Honesty when Honesty has Personal *and* Social Consequences

This study was virtually identical to Study S2, except that it included an additional condition (the "Consequences: Both" condition). In the "Consequences: Both" condition, the consequences of honesty impacted both the communicator and their partner.

In Part 1, we recruited communicators to make decisions. In Part 2, we recruited observers to judge communicators from Part 1.

2.3.1. Method for Study S3

Participants
- **Part 1 (pilot to generate Communicators, Study S3a)**
  - $N = 153$, 151 who provided demographics ($M_{age} = 33.68$, $SD_{age} = 9.78$; 97 males, 53 females, 1 prefer to self-describe)
  - Collected on MTurk
- **Part 2 (main study, Study S3b)**
  - $N = 598$ ($M_{age} = 36.19$, $SD_{age} = 10.64$; 304 males, 293 females, 1 prefer to self-describe)
  - Collected on MTurk

Procedure
- **Part 1 (pilot to generate Communicators; Study S3a)**
  - Participants were Communicators in the role of Decider.
  - Participants played a Coin Flip Game that was virtually identical to the Coin Flip Game our focal participants learned about in Study S2 (see Table S10), except we added a third condition in which the Communicator's reported outcome could impact both their partner's bonus and their own bonus.
  - Specifically, participants were assigned to one of three consequences conditions (Consequences: Personal vs. Social vs. Both) where they learned that their reported outcome impacted either their own bonus (Personal Consequences), the bonus of another person (Social Consequences), or both their own bonus and the bonus of another person (Both Consequences).
- **Part 2 (main study; Study S3b)**
  - Participants were observers who judged a Communicator who had previously participated as a Decider in the Coin Flip Game in Part 1.
  - Participants learned about their partner's behavior in the Coin Flip Game. Specifically, they learned that their partner was asked to report the outcome of a digital coin flip and were assigned in a 2 (Decision strategy: Unconditional Honesty vs. Looking) x 3 (Consequences: Personal vs. Social vs. Both) between-subjects design.
  - Participants then played the same Trust Game as Study S2 and judged the Communicator on a series of traits.

Dependent Variables
- **Part 1 (pilot to generate Communicators, Study S3a)**
  - Our primary dependent variable was the Communicator's report of the coin flip (Unconditional Honesty, Unconditional Lying, or Looking).
- **Part 2 (main study, Study S3b)**
  - Our main dependent variable was judgment of the Communicator's morality ($\alpha = .968$) using the same three items as prior studies.

- o We also measured the Communicator's trustworthiness (α = .940) and honesty using the same items as Study S2.
- o Benevolence was measured as a composite of the items kind, benevolent, and well-intended (α = .681).
- o We also measured integrity ($r(599)$ = .440) as a composite of the items principled and committed.
- o We also included an exploratory measure of the Communicator's thoughtlessness, measured using a three-item composite of thoughtful (reverse-scored), careless, and impulsive (1 = Not at all, 7 = Extremely; α = .448).
- o Behavioral trust was measured using the same Trust Game as Study S2.

Analyses
- **Part 1 (pilot to generate Communicators, Study S3a)**
  - o We ran a chi-squared test to examine Communicators' decisions across the Consequences conditions.
- **Part 2 (main study, Study S3b)**
  - o We ran a 2x3 ANOVA on our attitudinal variables (morality, trustworthiness, benevolence, integrity, honesty, and thoughtlessness) using Decision Strategy and Consequences as factors. We decomposed significant interactions using t-tests. Specifically, we examined judgments of Decision Strategy (Unconditional Honesty vs. Looking) for each of the Consequences conditions (Personal/Social/Both).
  - o We ran a logistic regression on trust in the Trust Game (1 = pass, 0 = keep) using Decision Strategy (1 = Unconditional Honesty, 0 = Looking), Consequences condition (Personal/Social/Both), and their interactions as independent variables.

2.3.2. Results & Discussion for Study S3

- **Part 1 (pilot to generate Communicators; Study S3a)**
  Communicator Decisions

    Communicators used similar decision strategies across all three consequences conditions (χ^2 = 4.15, p = .386), with the majority of participants choosing Unconditional Honesty. See Table S13 for percentages of each decision strategy in each condition.

    In Study S3a, Communicators were randomly assigned HEADS or TAILS, which allows us to separate whether telling the truth after looking would ultimately be helpful or harmful. Communicators who chose Looking typically made the prosocial decision, especially when the truth was helpful.

Table S13.

Communicators' Reported Decision Strategy in Study S3a

	Communicator's Initial Decision Strategy			Communicator's Decision Strategy After Looking			
	Un-conditional Honesty	Un-conditional Lying	Look	Truth Helps		Truth Harms	
				Truth	Lie	Truth	Lie
Personal	51.0%	9.8%	39.2%	90.9%	9.1%	12.5%	87.5%
Social	60.0%	8.0%	32.0%	100.0%	0.0%	33.3%	67.7%
Both	52.9%	2.0%	45.1%	100.0%	0.0%	14.3%	85.7%

Notes. In the Consequences: Both condition, Truth Helps refers to a truth that helps the self but harms others. Conversely, Truth Harms refers to a truth that harms the self but helps others.

- **Part 2 (main study, Study S3b)**
  Morality
    We found a main effect of Decision Strategy on judgments of morality ($F(1,592) = 101.97$, $p = .001$) such that those who were honest without looking at the consequences were judged to be more moral than those who looked at the consequences of honesty before making a decision. We found no main effect of Consequences ($p = .053$). However, we did find a Decision Strategy x Consequences interaction ($F(1,592) = 4.11$, $p = .017$) on judgments of morality; the effect of Looking was greater when the consequences were Personal or Personal *and Social*, rather than just Social, though the simple effects of unconditional honesty versus looking on morality were significant in all three conditions. Descriptive statistics are presented in Table S14.

  Other Measures
    We found a main effect of Decision Strategy on judgments of trustworthiness ($F(1,592) = 13.34$, $p < .001$), benevolence ($F(1,592) = 4.52$, $p = .034$), integrity ($F(1,592) = 5.24$, $p = .022$), and honesty ($F(1,592) = 10.60$, $p < .001$). Communicators who engaged in Unconditional Honesty were rated as more trustworthy, more benevolent, having more integrity, and more honest than Communicators who looked. We did not find a significant effect of Decision Strategy on judgments of thoughtlessness ($p = .252$).
    We did not find an effect of Consequences on judgments of trustworthiness, benevolence, integrity, honesty, or thoughtlessness ($ps \geq .313$). We also did not find a Decision Strategy x Consequences interaction for trustworthiness, benevolence, integrity, or honesty ($ps \geq .630$). However, there was a significant Decision Strategy x Consequences interaction effect for judgments of thoughtlessness ($F(2,592) = 4.10$, $p = .017$). Descriptive statistics are presented in Table S14.

Table S14.

Descriptive Statistics for other Measures in Study S3b

Dependent Variables	Personal		Social		Both	
	Uncon. Honesty	Looking	Uncon. Honesty	Looking	Uncon. Honesty	Looking
Morality	5.41 (1.13)	4.12 (1.56)	5.01 (1.25)	4.33 (1.31)	5.14 (1.37)	3.74 (1.49)
Trustworthiness	3.79 (1.71)	3.13 (1.69)	3.68 (1.65)	3.35 (1.68)	3.75 (1.74)	3.23 (1.59)
Benevolence	3.98 (1.17)	3.83 (1.57)	4.10 (1.36)	3.70 (1.52)	3.85 (1.34)	3.63 (1.76)
Integrity	4.08 (1.19)	3.80 (1.56)	4.17 (1.35)	3.87 (1.53)	3.93 (1.35)	3.70 (1.67)
Honest	4.63 (1.45)	4.09 (2.17)	4.29 (1.32)	4.02 (1.94)	4.41 (1.44)	3.79 (2.19)
Thoughtlessness	4.07 (1.78)	3.67 (1.66)	3.75 (1.65)	4.08 (1.83)	3.62 (1.86)	4.19 (1.77)
Trust Behavior	41.2%	28.8%	39.4%	41.6%	41.2%	40.2%

Note. The descriptives reported here come from a MANOVA with all dependent variables, and therefore reflect descriptives only from the set of participants who responded to ALL variables. In the meta-analysis reported in the main manuscript, we report descriptives of morality, based on all participants who responded to that scale only, which leads to slight changes in overall descriptives.

In the Trust Game, there were no significant effects of Decision Strategy, Social Consequences or Both Consequences, nor were there significant interactions. The regression results from the Trust Game are displayed in Table S15.

Table S15.

Regression Results on Passing Behavior in the Trust Game in Study S3b

	b	*SE*	*p*
Constant	-0.90	0.22	.000
Decision Strategy	-0.55	0.30	.065
Social Consequences	-0.56	0.30	.057
Consequences for Both	-0.51	0.30	.091
Decision Strategy * Social Consequences	0.64	0.41	.123
Decision Strategy * Consequences for Both	0.50	0.42	.226

Notes. The displayed results are from a logistic regression using passing behavior as our dichotomous DV (1 = pass, 0 = keep) and Decision Strategy (1 = Unconditional Honesty, 0 = Looking), Consequences for Others (1 = Social Consequences, 0 = all other conditions), Consequences for Both (1 = Both Consequences, 0 = all other conditions), and all of their interactions as independent variables.

Discussion
In Study S3a, communicators made similar choices whether honesty had consequences for themselves, others, or both themselves and others. This suggests that the decision to be honest without looking is not influenced by who is impacted by the consequences of honesty. Study S3a also suggests that unconditional honesty is the most common decision strategy.

In Study S3b, we replicate and extend our results from Study S2: people judged communicators who engaged in unconditional honesty as more moral, trustworthy, benevolent, honest, and higher in integrity than communicators who looked. Although we found a stronger effect of unconditional honesty on moral judgments when the consequences of honesty were *Personal,* the benefits of unconditional honesty also persisted when the consequences were Social and both Personal and Social.

2.4. Study S4: Judgements of Unconditional Honesty and Harm Salience

In this study, we examined whether the preference for unconditional honesty was attenuated when the potential harm caused by honesty was more salient. As in Study S3, we first recruited participants as communicators to make real decisions in Part 1, and we recruited participants as observers to judge these communicators' decisions in Part 2.

2.4.1. Method for Study S4

Participants
- **Part 1 (pilot to generate Communicators; Study S4a)**
  - $N = 198$ ($M_{age} = 35.53$, $SD_{age} = 10.62$; 116 males, 81 females, 1 prefer to self-describe)
  - Collected on MTurk
- **Part 2 (main study; Study S4b)**
  - $N = 799$ ($M_{age} = 36.41$, $SD_{age} = 10.95$; 383 males, 414 females, 2 prefer to self-describe)
  - Collected on MTurk

Procedure
- **Part 1 (pilot to generate Communicators Study S4a)**
  - Participants in Part 1 were called Communicators. They played a Coin Flip Game in which they were assigned to the role of "Decider" and were paired with a partner (assigned to the role of "Receiver").
    - Communicators played a version of the Coin Flip Game, which was nearly identical to the game described in Studies S2 and S3, except we altered the potential payouts associated with the coin flip outcomes. In all conditions, the payouts only went to the Receiver (i.e., honesty always had social consequences), and the Receiver always started the game with $2.50.
    - In the Harm Salient condition, Communicators learned the values of $A and $B were between -$2.50 and +$2.50. We expected that a potential loss would make harm salient.
    - In the Harm Not-Salient condition, Communicators learned the values of $A and $B were between $0 and +$2.50.
- **Part 2 (main study, Study S4b)**
  - Participants in Part 2 saw real decisions that were made by Communicators in Part 1. They were then asked to evaluate the Communicators.
  - Participants were assigned in a 2 (Consequences: Harm Salient vs. Harm Not-Salient) x 2 (Decision Strategy: Unconditional Honesty vs. Looking) design.
    - Participants saw the exact Coin Flip Game instructions that were shown to Communicators in Part 1 and were shown a Communicator who either chose Unconditional Honesty or Looking.
  - After learning about how the Communicator behaved in the Coin Flip Game, participants played the Choice Game (the Trust Game used in Study S2 and Study S3b) with the Communicator and rated them on a series of traits.

Dependent Variables
- **Part 1 (pilot to generate Communicators, Study S4a)**
  - o Our primary dependent measure was the Communicator's decision strategy.
  - o We also examined the ultimate report Communicators made if they chose Looking.
- **Part 2 (main study, Study S4b)**
  - o Our primary dependent measure was judgments of the Communicator's morality using the same three items as prior studies (α = .963).
  - o We also examined judgments of the Communicator's trustworthiness (α = .927), benevolence (r = .738), integrity (α = .918), and honesty using the same items as and Study S2.
  - o We also included the same behavioral measure of trust as in Study S2 and Study S3b: the decision to pass in the Choice Game (Trust Game).

Analyses
- **Part 1 (pilot to generate Communicators, Study S4a)**
  - o We ran a chi-squared test to examine the Communicators' decisions in the Harm Salient and Harm Not-Salient conditions.
- **Part 2 (main study, Study S4b)**
  - o We ran a 2x2 ANOVA to examine judgments of morality, trustworthiness, benevolence, integrity, and honesty using Consequences (Harm Salient vs. Harm Not-Salient), Decision Strategy (Unconditional Honesty vs. Looking), and their interaction as IV's.
  - o To examine passing behavior in the Choice Game, we ran a logistic regression (Pass = 1; Keep = 0) using Decision Strategy (1 = Unconditional Honesty; 0 = Looking) and Harm Consequences (1 = Harm Salient; 0 = Harm Not-Salient) as independent variables.
  - o We also ran a chi-squared test to examine the rates of passing behavior within each Consequences condition.

2.4.2. Results & Discussion for Study S4

- **Part 1 (pilot to generate Communicators, Study S4a)**
  Communicator Decisions
  In the Harm Not-Salient condition, 66.3% of Communicators chose Unconditional Honesty while 33.7% of Communicators chose Looking. In contrast, for the Harm Salient condition, 51.9% of Communicators chose Unconditional Honesty, 47.2% of Communicators chose Looking, and 0.9% chose Unconditional Lying (χ^2 = 4.80, p = .091).
  Of the Communicators who chose Looking in the Harm Not-Salient condition, 93.9% ultimately decided to report the honest outcome of the coin flip. All the Communicators who chose Looking in the Harm Salient condition ultimately decided to report the honest outcome of the coin flip after learning the values associated with the outcome. Notably, we always revealed that the honest outcome was the prosocial outcome to those who engaged in Looking (HEADS = $1.00, TAILS = $0.50).

- **Part 2 (main study, Study S4b)**
  Morality
  We found a significant effect of Decision Strategy on judgments of morality such that Communicators who chose Unconditional Honesty were judged as more moral than Communicators who chose Looking ($F(1,797)$ = 16.91, p < .001). We did not observe an effect of Consequences nor a Decision Strategy x Consequences interaction on judgments of morality ($ps \geq$ .169). Descriptive statistics are presented in Table S16.

Other Measures

We found a significant effect of Decision Strategy on judgments of benevolence ($F(1,797) = 6.57$, $p = .011$), integrity ($F(1,796) = 21.82$, $p < .001$), and honesty ($F(1, 797) = 50.12$, $p < .001$) such that Communicators who engaged in Unconditional Honesty were rated as more benevolent, having more integrity, and more honest than Communicators who chose Looking. We did not find an effect of Decision Strategy on perceptions of the Communicator's trustworthiness ($p = .147$).

We did not find any significant effects of Consequences on judgments (trustworthiness, benevolence, integrity, honesty: $ps \geq .077$). We did not find significant interactions of Decision Strategy x Consequences on judgments of trustworthiness, benevolence, integrity, and honesty ($ps \geq .138$). Descriptive statistics are presented in Table S16.

Table S16.

Descriptive Statistics for Attitudinal Measures in Study S4b

Dependent Variables	Harm Salient		Harm Not-Salient	
	Unconditional Honesty	Looking	Unconditional Honesty	Looking
Morality	4.73 (1.03)	4.42 (1.11)	4.63 (1.12)	4.29 (1.24)
Trustworthiness	4.05 (1.56)	3.97 (1.58)	3.93 (1.53)	3.69 (1.62)
Benevolence	4.36 (1.12)	4.23 (1.18)	4.34 (1.18)	4.04 (1.22)
Integrity	4.70 (0.98)	4.36 (1.14)	4.61 (1.08)	4.22 (1.19)
Honesty	4.86 (1.21)	4.36 (1.31)	4.91 (1.20)	4.14 (1.35)
Trust Behavior	58.3%	58.9%	55.7%	50.2%

Note. The descriptives reported here come from a MANOVA with all dependent variables, and therefore reflect descriptives only from the set of participants who responded to ALL variables.

Trust Game Results

In the Trust Game, results from a logistic regression indicated that participants were directionally more likely to pass to Communicators when harm was salient (vs. not; $p = .078$), though there were no effects significant at the $p < .05$ thresholds in this model. See Table S16 for the descriptive results of passing behavior in the Trust Game, and Table S17 for the regression results of passing behavior in the Trust Game.

Table S17.

Logistic Regression Results on Passing Behavior in the Trust Game in Study S4b

	b	*SE*	*p*
Constant	0.01	0.14	.944
Decision Strategy	0.22	0.20	.275
Harm Salience	0.35	0.20	.078
Decision Strategy * Harm Salience	-0.25	0.29	.387

Note. Regression model predicts passing behavior in the Trust Game (Pass = 1; Keep = 0) using Decision Strategy (1 = Unconditional Honesty; 0 = Looking) and Harm Consequences (1 = Harm Salient; 0 = Not Salient) as independent variables.

Discussion

In Study S4a, communicators were directionally, but not significantly more likely to choose Looking when harm was salient.

In Study S4b, we found that communicators who were honest without looking at the consequences of their decision (Unconditional Honesty) were judged to be more moral, benevolent, honest, and higher in integrity, but equally trustworthy compared to communicators who chose Looking, regardless of harm salience.

2.5. Study S5: Judgments of Unconditional Honesty and Perceptions of Selfishness

In Study S5, we consider whether the decision to look for information about the social consequences of honesty is penalized because it signals moral flexibility that could also be used to pursue one's own interests (i.e., that someone will also look at the *personal* consequences of honesty before deciding whether to tell the truth). We test this potential mechanism by orthogonally manipulating whether a communicator engages in unconditional honesty (vs. looking) when honesty influences others' outcomes, and whether a communicator engages in unconditional honesty (vs. looking) when honesty influences one's own outcomes.

2.5.1. Method for Study S5

Participants
- N = 540 participants (M_{age} = 35.83, SD_{age} = 11.01; 286 males, 251 females, 3 prefer to self-describe)
- Collected on MTurk

Procedure and materials
- In Study S5, participants learned about a Communicator who had played a game called the Number Game (see exact experimental stimuli in Table S18).
- Participants were randomly matched with a Communicator who played two rounds of the Number Game. In the first round of the game, the Communicator had to report the outcome of a random number generator, knowing that their report could either help or harm a target. Participants learned the Communicator was either honest without looking at the *social* consequences of their decision (Unconditional Honesty) or looked at the social consequences (Looking) before making a communication decision.
- Participants then learned that the Communicator played a second round of the Number Game, in which their report could either help or harm themselves (the Communicator). Participants learned that: 1) the Communicator was honest without looking at the *personal* consequences of their decision (Unconditional Honesty), 2) the Communicator looked at the personal consequences (Looking) before making a communication decision, or 3) the Communicator's decision in this round was unknown.
- In sum, participants were randomly assigned to a condition in a 2 (Decision for Other/Round 1: Looking vs. Unconditional Honesty) x 3 (Decision for Self/Round 2: Looking vs. Unconditional Honesty. vs. No Info) between-subjects design.

Dependent Variables
- Our primary dependent variable was participants' judgments of the Communicator's morality using the same three-item composite as past studies (α = .924).
- Our secondary dependent variable was the participant's belief about whether the Communicator would choose to look at the personal consequences of truth-telling (i.e., "How likely is it that the Communicator chose to find out the values of X and Y before making a decision?"; anchored at 0% and 100%). This served as a manipulation check in the "Looking for Self" and "Unconditional Honesty for Self" conditions. However, this item also allowed us to test whether Looking at the *social* consequences of honesty is perceived as a signal of one's propensity of Looking at the *personal* consequences of honesty in the "No Info for Self" condition. For completeness, we also measured how likely participants believed it was that the Communicator chose to report an odd number and that the Communicator chose to report an even number.

- Participants also reported their judgments of the Communicator's general lying tendencies. Participants indicated the extent to which they believed the Communicator would do the following things (1 = Strongly disagree, 7 = Strongly agree): lie to help others, lie in arbitrary situations, lie across situations, lie to hurt others, and lie to help themselves.

Table S18.

Experimental Stimuli from Study S5 (Number Game Instructions)

Number Game Round 1: Decision for Other	Number Game Round 2: Decision for Self
In the **NUMBER GAME,** your partner was in the role of **DECIDER**, and started the game with $0. Your partner was paired with a different person (not you), who was in the role of **RECEIVER.** The **RECEIVER** started the game with $0. On the next page are the exact instructions your partner saw as **DECIDER**.	Here your partner was in the role of Decider. They played another round of the Number Game, except in this round, the values of X and Y were payoffs for them, the Decider. That is, their decisions in the game affected how much they, rather than this partner, earned. Here are the instructions your partner saw as the Decider.

NUMBER GAME-DECIDER INSTRUCTIONS
You will get to choose an allocation for your partner, the RECEIVER. The allocation will depend on your choice and the outcome of a random number generator.
 	
Specifically, you will have to report the outcome of the random number generator.
 	
The random number generator will generate a number between 1 and 9 (inclusive) and will be either 1, 2, 3, 4, 5, 6, 7, 8, or 9.
 	
Then, you will have to report whether the number generated was ODD or EVEN.
 	
If you report that the number is ODD, your partner (the RECEIVER) will be paid according to Option A.
 	
If you report that the number is EVEN, your partner (the RECEIVER) will be paid according to Option B.
    • Option A) $X for the RECEIVER
    • Option B) $Y for the RECEIVER
 	
The values of X and Y are between $0 and $10, but you do not know their exact values. You can
	NUMBER GAME-DECIDER INSTRUCTIONS
	You will get to choose an allocation for you, the DECIDER. The allocation will depend on your choice and the outcome of a random number generator.
	
	Specifically, you will have to report the outcome of the random number generator.
	
	The random number generator will generate a number between 1 and 9 (inclusive) and will be either 1, 2, 3, 4, 5, 6, 7, 8, or 9.
	
	Then, you will have to report whether the number generated was ODD or EVEN.
	
	If you report that the number is ODD, you (the DECIDER) will be paid according to Option A.
	
	If you report that the number is EVEN, you (the DECIDER) will be paid according to Option B.
	    • Option A) $X for the DECIDER
	    • Option B) $Y for the DECIDER
	
	The values of X and Y are between $0 and $10, but you do not know their exact values. You can choose to reveal their exact values before making your decision.

choose to reveal their exact values before making your decision.	
A random number has been generated. The number is: 7 Please remember this number. Feel free to write it down.	A random number has been generated. The number is: 1 Please remember this number. Feel free to write it down.
Seeing that random number generator produced an **ODD** number, here is how your partner responded. *[Presented in Unconditional Honesty Condition]* Please make a decision. This decision will affect your partner. **The number generated was ODD** *This will lead to Option A) $X for the Receiver* The number generated was EVEN *This will lead to Option B) $Y for the Receiver* I choose not to report a number. *I would like to find out the values of X and Y and then make a decision.* *[Presented in Looking Condition]* Please make a decision. This decision will affect your partner. The number generated was ODD *This will lead to Option A) $X for the Receiver* The number generated was EVEN *This will lead to Option B) $Y for the Receiver* **I choose not to report a number.** *I would like to find out the values of X and Y and then make a decision.*	Seeing that random number generator produced an **ODD** number, here is how your partner responded. *[Presented in Unconditional Honesty Condition]* Please make a decision. This decision will affect only you. The number generated was ODD *This will lead to Option A) $X for you, the Decider* The number generated was EVEN *This will lead to Option B) $Y for you, the Decider* **I choose not to report a number.** *I would like to find out the values of X and Y and then make a decision.* *[Presented in Looking Condition]* Please make a decision. This decision will affect only you. The number generated was ODD *This will lead to Option A) $X for you, the Decider* The number generated was EVEN *This will lead to Option B) $Y for you, the Decider* **I choose not to report a number.** *I would like to find out the values of X and Y and then make a decision.* *[No screenshot in No Info Condition]*
Now you will see another decision that your partner made in the Number Game.	

Analyses
- We ran a 2x3 ANOVA on judgments of morality, probability of looking for self, and general lying tendencies using Decision for Other and Decision for Self as factors.

2.5.2. Results & Discussion for Study S5

Morality

A 2x3 ANOVA on judgments of morality revealed a main effect of Decision for Other ($F(1,534)$ = 22.27, $p < .001$) such that participants judged Communicators who engaged in Unconditional Honesty ($M = 5.07$, $SD = 1.15$) as more moral than Communicators who engaged in Looking ($M = 4.60$, $SD = 1.20$).

We also found a main effect of Decision for Self on judgments of morality ($F(2,534) = 11.52$, $p < .001$). Communicators who engaged in Unconditional Honesty for themselves ($M = 5.09$, $SD = 1.21$) were judged as more moral than Communicators who engaged in Looking for themselves ($M = 4.52$, $SD = 1.21$; $p < .001$). Communicators in the No Info condition ($M = 4.90$, $SD = 1.11$) were also judged as more moral than those who engaged in Looking for themselves ($p = .002$), but no different than

Communicators who engaged in Unconditional Honesty ($p = .119$). We did not find a significant Decision for Other x Decision for Self interaction on morality ($F(2,534) = 1.80$, $p = .166$).

Probability of Looking for Self

A 2x3 ANOVA revealed a main effect of Decision for Self on the perceived probability of looking for self ($F(2,534) = 87.28$, $p < .001$). Consistent with the intent of our manipulation, Communicators who chose to look at the personal consequences of their honesty ($M = 63.04$, $SD = 35.77$) were rated as more likely to look at the personal consequences of their honesty than Communicators who engaged in Unconditional Honesty ($M = 21.81$, $SD = 24.06$, $p < .001$) and Communicator's whose decisions were unknown (No Info condition: $M = 39.74$, $SD = 33.18$, $p < .001$). Moreover, Communicators who engaged in Unconditional Honesty were perceived as less likely to look at the personal consequences than Communicators whose decision was unknown in the No info condition ($p < .001$).

We also found a main effect of Decision for Other ($F(1,534) = 67.95$, $p < .001$). Relative to Communicators who engaged in Unconditional Honesty in the social consequences condition ($M = 31.09$, $SD = 31.52$), Communicators who looked at the social consequences of their honesty ($M = 52.19$, $SD = 36.47$) were rated as more likely to look at the personal consequences of honesty. In other words, Looking at the social consequences of honesty was indeed seen as a signal that one would also look at the personal consequences of honesty. We did not find a significant Decision for Other x Decision for Self interaction ($p = .369$).

Taken together, the morality and probability-of-looking-for-self measures suggest that Looking at the social consequences of honesty is perceived as a signal of a Communicator's propensity to look at the *personal* consequences of honesty. However, this perception does not seem to be the main driver of the moral preference for unconditional honesty. Communicators who looked at the social consequences of honesty were still penalized, even when it was known that they committed to honesty when it could be costly for themselves.

General Lying Tendencies

We ran a 2x3 ANOVA examining perceptions of the Communicator's general lying tendencies using Decision for Self and Decision for Other as factors. This analysis revealed a main effect of Decision for Other such that those who looked at the social consequences of honesty were rated as more likely to lie to help others ($F(1,531) = 15.25$ $p < .001$), lie in arbitrary situations ($F(1,531) = 15.25$, $p < .001$), lie across situations ($F(1,531) = 19.96$, $p < .001$), lie to hurt others ($F(1,504) = 18.12$, $p < .001$), and lie to help themselves ($F(1,530) = 22.75$, $p < .001$) than those who chose Unconditional Honesty. In other words, Looking, relative to Unconditional Honesty, signaled that a Communicator was more likely to tell all types of lies.

We also found a main effect of Decision for Self, such that those who engaged in Looking at the personal consequences of honesty were seen as more likely than those in the control condition and those who engaged in Unconditional Honesty for Self to lie in arbitrary situations ($F(2,531) = 4.53$, $p = .011$), lie across situations ($F(2,531) = 4.53$, $p = .011$), and lie to help themselves ($F(2, 531) = 7.35$, $p = .001$). There was no main effect for the items "lie to help others" and "lie to hurt others" ($ps > .591$).

Finally, we found a significant Decision for Other x Decision for Self interaction ($F(1,531) = 5.27$, $p = .005$) on lying in arbitrary situations: Looking at consequences for the self signaled greater lying relative to Unconditional Honesty for the Self and the Control condition when the Communicator engaged in Unconditional Honesty for Others. However, Looking for the self, Unconditional Honesty for the Self, and the Control were rated no differently from each other when the Communicator engaged in Looking for Others. There was no significant interaction for any of the other general lying dependent variables ($ps > .115$). Descriptive statistics are presented in Table S19.

Table S19.

Descriptive Statistics of Other Measures in Study S5

Decision for Self	Decision for Other (Unconditional Honesty)			Decision for Other (Looking)		
	Unconditional Honesty	Looking	Control	Unconditional Honesty	Looking	Control
Lie to help others	3.48 (1.64)	3.75 (1.51)	3.75 (1.63)	4.24 (1.38)	4.04 (1.53)	4.24 (1.44)
Lie in arbitrary situations	3.19 (1.81)	4.13 (1.42)	3.35 (1.55)	4.14 (1.42)	4.13 (1.56)	4.20 (1.48)
Lie across situations	3.17 (1.84)	3.92 (1.37)	3.45 (1.40)	4.01 (1.47)	4.22 (1.58)	4.09 (1.52)
Lie to hurt others	2.88 (1.69)	3.18 (1.39)	3.01 (1.58)	3.55 (1.61)	3.57 (1.61)	3.68 (1.50)
Lie to help themselves	3.38 (1.67)	4.32 (1.56)	4.13 (1.51)	4.42 (1.64)	4.68 (1.57)	4.68 (1.50)

Discussion

Study S5 suggests that beliefs about a communicator's likelihood of *selfishly* engaging in selective honesty do not drive the ethical preference for unconditional honesty.

2.6. Study S6: Unconditional Prosociality

In Study S6, we begin to examine whether the preference for unconditional moral rule-following extends beyond honesty. We manipulate whether communicators start with information about what is honest (and have to decide whether to find out whether their honesty would help or harm others before making a communication decision) or with information about what is helpful (and have to decide whether 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.

As in Studies S3 and S4, we recruited Communicators to actually make decisions in Study S6 – Part 1 (Study S6a). We used these Communicator decisions as stimuli in Study S6 – Part 2 (Study S6b) and Study S7.

2.6.1. Method for Study S6

Participants
- **Part 1 (pilot to generate Communicators, Study S6a)**
  - $N = 200$ ($M_{age} = 35.57$, $SD_{age} = 10.55$; 102 males, 98 females)
  - Collected on MTurk
- **Part 2 (main study, Study S6b)**
  - $N = 798$ ($M_{age} = 37.52$, $SD_{age} = 11.79$; 383 males, 415 females)
  - Collected on MTurk

Procedure
- **Part 1 (pilot to generate Communicators, Study S6a)**
  - Participants played a variant of the Number Game and were assigned to one of two Context conditions (Honesty vs. Prosociality). Specifically, participants were asked to report the outcome (odd/even) of a random number generator.
    - In the Honesty Condition, participants knew the true outcome of the number generator but did not know if the outcome would be helpful or harmful to another person (i.e., they did not know whether reporting the truth or a lie would result in another person gaining or losing a dollar).
    - In the Prosociality Condition, participants knew whether reporting odd or even would result in another person gaining or losing a dollar but did not know what number was actually produced by the random number generator.
  - Participants then made the decision to either report the outcome of the random number generator (odd/even) or find out full information before reporting.
- **Part 2 (main study, Study S6b)**
  - Participants were assigned to a Communicator who had recently played a variant of the Number Game, similar to Study S5.
  - Participants learned that the Communicator had to report the outcome of a random number generator (odd/even) and were randomly assigned to a condition in a 2 (Context: Honesty vs Prosociality) x 2 (Decision Strategy: Looking vs Unconditional Behavior) design.
  - Participants then played the same Trust Game as in Studies S2-S4 with the Communicator, and made judgments of the Communicator's morality, trustworthiness, benevolence, integrity, and honesty.

Dependent Variables
- **Part 1 (pilot to generate Communicators Study S6a)**

- o Our primary dependent variable was the Communicator's decision strategy (Unconditional Behavior or Looking, i.e., Unconditional Honesty vs Looking in Honesty context; Unconditional Prosociality vs Looking in Prosociality context)
- **Part 2 (main study, Study S6b)**
  - o Our primary dependent measure was judgments of the Communicator's morality (α = .966) using the same three items as previous studies.
  - o We also examined judgments of the Communicator's trustworthiness (α = .925), benevolence (α = .873), integrity (α = .929), and honesty using the same items as Studies S2 and S4b
  - o Trusting behavior was measured using the same Trust Game (The Choice Game) as S2, S3b, and S4b

Analyses
- **Part 1 (pilot to generate Communicators, Study S6a)**
  - o We ran a chi-squared test to examine the proportion of people in each context who chose Unconditional Behavior or Looking.
- **Part 2 (main study, Study S6b)**
  - o We ran a 2x2 ANOVA on judgments of morality, trust, benevolence, integrity, and honesty using Context and Decision Strategy as factors. We also ran planned contrasts to decompose significant interactions.
  - o We also ran a logistic regression examining trust behavior using Trust (1 = Pass $2, 0 = Keep $2), Context (1 = Prosociality, 0 = Honesty) and Decision Strategy (1 = Unconditional Behavior, 0 = Looking) as factors. Chi-squared tests were conducted to examine significant effects.

2.6.2. Results & Discussion for Study S6
- **Part 1 (pilot to generate Communicators, Study S6a)**
  Communicator Decisions
  In Part 1, we found that Communicators made significantly different initial decisions in the Honesty Context and the Prosociality Context (χ^2 = 7.25, p = .027). In the Honesty Context, 62.0% chose Unconditional Honesty, 11.0% chose Unconditional Lying, and 27.0% chose Looking. In contrast, 43.0% chose Unconditional Prosociality, 17.0% chose Unconditional Harm (antisociality), and 40.0% chose Looking in the Prosociality Context. Across Context conditions, we always revealed that the honest outcome was also the prosocial outcome to those who engaged in Looking. Of the Communicators who chose Looking, 96.3% chose to tell the helpful truth in the Honesty Context and 95.0% chose to tell the helpful truth in the Prosocial Context. This suggests that Unconditional behavior (vs. Looking) is preferred as a decision strategy across conditions, but that Communicators display a stronger preference for Unconditional Honesty than Unconditional Prosociality.

- **Part 2 (main study, Study S6b)**
  Morality
  In Part 2, we found a significant effect of Context on morality *(F(1,795) = 7.87, p = .005)* where Communicators in the Prosociality condition were rated as more moral than those in the Honesty condition.
  We also found a significant effect of Decision Strategy on morality *(F(1,795) = 14.76, p < .001)* where Communicators who chose Unconditional behavior were rated as more moral than Communicators who engaged in Looking.
  Importantly, this was qualified by a significant Context x Decision Strategy interaction *(F(1,795) = 12.44, p < .001)* such that participants judged Communicators who did not look for

more information as more moral than those who engaged in Looking in the Honesty Context ($p <$ .001) but not in the Prosociality context ($p = .805$). In other words, observers, like Communicators, display a preference for Unconditional Honesty but not necessarily Unconditional Prosociality. Descriptive statistics for morality are presented in Table S20.

Other Measures

We found significant effects of Context on ratings of benevolence and integrity where Communicators in the Prosociality condition were rated as more benevolent ($F(1,795) = 8.69$, $p =$ .003) and higher in integrity ($F(1,795) = 10.86$, $p = .001$) than those in the Honesty condition. There effects of Context for judgments of trustworthiness and honesty were not significant at the $p < .05$ thresholds, but directionally those in the Prosociality condition were rated as more trustworthy ($F(1,795) = 3.43$, $p = .064$) and more honest ($F(1,795) = 3.44$, $p = .064$) than those in the Honesty condition.

Additionally, we found a significant effect of Decision Strategy such that Communicators who chose Unconditional behavior were viewed as more trustworthy than Communicators who chose Looking ($F(1,795) = 6.90$, $p = .009$), benevolence ($F(1,795) = 24.81$, $p < .001$), were rated higher in integrity ($F(1,795) = 5.58$, $p = .018$), and as more honest ($F(1,795) = 11.28$, $p < .001$).

Finally, we found significant Context x Decision Strategy interactions for ratings of integrity and honesty ($Fs \geq 11.41$, $ps \leq .001$) such that Communicators who chose Unconditional behavior were afforded higher ratings of integrity ($p < .001$) and honesty ($p < .001$) than Communicators who chose Looking in the Unconditional Honesty Context, but not in the Unconditional Prosociality Context ($ps > .092$). We did not find significant Context x Decision Strategy interactions for ratings of trust or benevolence ($ps \geq .121$). Descriptive statistics for trustworthiness, benevolence, integrity, and honesty are presented in Table S20.

Table S20.

Descriptive Statistics for Other Measures in Study S6b

Dependent Variables	Honesty		Prosociality	
	Unconditional honesty	Looking	Unconditional Prosociality	Looking
Morality	4.80 (1.06)	4.19 (1.27)	4.74 (1.16)	4.71 (1.20)
Trustworthiness	4.19 (1.47)	3.73 (1.62)	4.22 (1.57)	4.10 (1.58)
Benevolence	4.60 (1.12)	4.14 (1.25)	4.83 (1.26)	4.42 (1.29)
Integrity	4.64 (1.12)	4.16 (1.28)	4.63 (1.14)	4.72 (1.18)
Honesty	5.02 (1.08)	4.23 (1.35)	4.68 (1.21)	4.89 (1.19)
Trusting Behavior	53.3%	53.0%	59.6%	59.9%

Trust Game Results

A logistic regression revealed no effect of Decision Strategy, Context, or a Decision Strategy x Context interaction (*ps* ≥ .163). Regression results are presented in Table S21 and descriptive statistics of passing behavior in the Trust Game (i.e., trusting behavior) are presented in Table S20.

Table S21.

Regression Results on Passing Behavior in the Trust Game in Study S6b

	b	*SE*	*p*
Constant	0.40	0.14	.005
Decision Strategy	-0.01	0.20	.950
Context	-0.28	0.20	.163
Decision Strategy * Context	0.02	0.29	.935

Notes. Regression model predicts passing behavior in the Trust Game (pass = 1; keep = 0) using Decision Strategy (1 = Unconditional Behavior; 0 = Looking) and Context (1 = Prosociality; 0 = Honesty) as independent variables.

Discussion

Overall, this study reveals that both communicators and observers seem to prefer Unconditional Honesty (over Looking), but not Unconditional Prosociality (over Looking). Observers judged Unconditional Honesty to be more moral and as signaling higher integrity than Looking, but these effects were muted in the Prosociality context. However, participants judged unconditional moral behaviors as signaling higher trust and benevolence, regardless of context.

2.7. Study S7: Unconditional Prosociality 2

In Study S7, we provide a second test of the prediction that the preference for Unconditional Honesty is attenuated when it is certain that Looking does not lead to harmful lies (as in Study 5 of the main manuscript). We test this by examining perceptions of Unconditional Honesty and Looking once the outcomes of those decision strategies are known. We also provide a second test of whether the preference for unconditional moral rule-following extends beyond honesty and into the domain of prosociality (as in Study S6).

2.7.1. Method for Study S7

Participants
- $N = 799$ ($M_{age} = 37.44$, $SD_{age} = 12.01$; 377 males, 422 females)
- Collected on MTurk

Procedure
- Participants learned about a Communicator who had previously participated in a study in which they played "the Number Game," which was similar to the version used in Study S5 (see Table S18). In this game, Communicators were asked to report whether a number generated by a random number generator was odd or even. The Communicator's report was associated with monetary consequences for a partner, called the Target.
  - In the Honesty context, the Communicator first learned the true outcome of the random number generator, without knowing how their report would affect their partner (i.e., the Communicator knew the true outcome of the number generator but did not know how much money their partner would earn if they reported that the number was odd (versus even). This is conceptually identical to the coin flip context we examined in Studies 2, 3 and 5.
  - In the Prosociality Context, the Communicator first learned how their report of the number would affect their partner (i.e., the Communicator knew the monetary values associated with reporting odd versus even) but did not know the true outcome of the random number generator. Therefore, Communicators in this context could choose to report something that was ultimately prosocial (or antisocial), without knowing whether their report was true or false.
- Participants first learned about the Number Game. Then, participants learned the Communicator's decision strategy.
  - In the Unconditional Behavior conditions, the Communicator chose to initially report a number without full information (either about how the report would affect the Target or about the true value of the report). Notably, in the Honesty Context/Truth Outcome Condition, Unconditional Behavior is identical to our operationalization of Unconditional Honesty in prior studies. After learning of the Communicator's decision strategy, participants learned the consequences of this decision (i.e., whether the decision ultimately led to a helpful, harmful, honest, or dishonest report).
  - In the Looking conditions, the Communicator chose to look and learn about the unknown information before reporting a number. In the Honesty Context, Communicators who looked found out the monetary values associated with reporting odd versus even and therefore whether truth telling would help or harm the Target. Then, the Communicator either helped or harmed the Target, via truth telling or lying (by reporting odd or even). In the Prosociality Context, Communicators who looked found out whether the number generated was actually odd or even, and therefore whether the message that helped the

Target was honest or deceptive. Then the Communicator either helped or harmed the Target, via truth telling or lying.

- Regardless of context, we informed participants of the ultimate choice the Communicator made in each context, including whether the Communicator chose to look at the consequences of the report, if the final decision was honest or deceptive, and whether it helped or harmed the Target. That is, participants in this study were randomly assigned to a condition from a 2 (Context: Honesty vs. Prosociality) x 2 (Decision Strategy: Unconditional Behavior vs. Looking) x 2 (Honesty Outcome: Truth vs. Lie) x 2 (Prosocial Outcome: Help vs. Harm) between-subjects design. The full design is depicted in Figure S1.

Dependent Variables
- The main dependent variable was moral judgment of the Communicator; we used the same three-item composite as previous studies ($\alpha = .971$).
- We measured trustworthiness ($\alpha = .947$), benevolence ($\alpha = .906$), integrity ($\alpha = .903$), and honesty using the same items as Studies S2, S4b, and S6b.
- Trusting behavior was measured using the same Trust Game (The Choice Game) as Studies S2, S3b, S4b, and S6b.

Figure S1.

Design of the Number Game in Study S7

Note. This figure depicts the design of the Number Game. Participants learned about a Communicator and their decision in a between-subjects design. That is, they learned about one of the sixteen possible outcomes listed in the final column. The "Target" was referred to as the Communicator's "partner" within the Number Game instructions.

Analyses

- We preregistered and ran a 2 (Context: Honesty vs. Prosociality) x 2 (Decision Strategy: Unconditional Behavior vs. Looking) x 2 (Honesty Outcome: Truth vs. Lie) x 2 (Prosocial Outcome: Help vs. Harm) ANOVA on judgments of morality, trust, benevolence, integrity, and honesty using Context and Decision Strategy as factors. We report those results in Table S23.
- However, below, we focus on the contrasts that test our theoretical questions of interest regarding judgments of moral character.

2.7.2 Results & Discussion for Study S7

- ***Is Unconditional Honesty seen as a more moral decision strategy than Looking, holding outcomes constant?***
  To examine this question, we compare moral judgments of Unconditional Honesty (cells B and D in Figure S2, Panel A) to Looking, holding outcomes constant (cells F and H in Figure S2, Panel A). Specifically, we examine the effect of Decision Strategy (Unconditional Behavior vs. Looking) and Prosocial Outcome (Help vs. Harm) among Communicators who told the truth (Honesty Outcome: Truth) in the Honesty Context conditions.
  We find a main effect of Decision Strategy ($F(1,197) = 5.62$, $p = .019$) such that Unconditional Honesty ($M = 4.79$, $SD = 0.96$) is seen as more moral than Looking ($M = 4.40$, $SD = 1.43$). We also find a main effect of Prosocial Outcome ($F(1,197) = 14.07$, $p < .001$) such that Help ($M = 4.91$, $SD = 0.96$) is seen as more moral than Harm ($M = 4.25$, $SD = 1.41$).
  Importantly, these main effects were qualified by a significant interaction between Decision Strategy and Prosocial Outcome ($F(1,197) = 11.65$, $p < .001$). Unconditional Honesty ($M = 4.76$, $SD = 1.04$) was seen as more ethical than Looking ($M = 3.81$, $SD = 1.55$) when these decision strategies led to Harm ($< .001$), but Unconditional Honesty ($M = 4.82$, $SD = 0.88$) was seen similarly as Looking ($M = 4.99$, $SD = 1.03$) when these decisions led to Help ($p = .371$). Put differently, harmful truths were seen as more ethical when they resulted from unconditional behavior than when they reflected the choice a Communicator made after looking. However, prosocial truths were seen as equally ethical, regardless of the strategy the Communicator used.
- ***Is Unconditional Honesty seen as a more moral decision strategy than Looking, when Looking does not result in harmful lies?***
  We also compare moral judgments of Unconditional Honesty (cells B and D in Figure S2, Panel A) to Looking when it results in prosociality (cells F and G in Figure S2, Panel A). In other words, we compare the strategy of Unconditional Honesty to the strategy of Looking and avoiding harm.
  To do this, we conduct a simple t-test to compare the moral judgment of Unconditional Honesty (Context: Honesty, Decision Strategy: Unconditional Behavior, Honesty Outcome: Truth) to the moral judgment of Prosocial Looking (Context: Prosociality, Decision Strategy: Looking, Prosocial Outcome: Help). There was no difference between Unconditional Honesty ($M = 4.79$, $SD = 0.96$) and Looking ($M = 4.90$, $SD = 1.13$) when Looking resulted in prosocial outcomes ($t(196) = .716$, $p = .475$). Together, these two analyses corroborate the result of Study 5: People judge Unconditional Honesty (vs. Looking) as a more ethical decision strategy *unless* it is certain that looking will not lead to harm.
- ***In the Prosociality Context, is Unconditional Behavior seen as a more moral decision strategy than Looking, holding outcomes constant?***
  Next, we examine whether people also exhibit a preference for unconditional moral rule-following in the context of prosociality. We conduct a similar analysis to the first analysis we conducted in the Honesty Context: We compare moral judgments of Unconditional Prosociality (cells J and K in Figure S2, Panel B) to Looking (cells N and O in Figure S2, Panel B), holding outcomes

constant. In other words, we compare judgments of prosocial statements, when the truthfulness of those statements are or are not yet known.

To do this, we examine the effect of Decision Strategy (Unconditional Behavior vs. Looking) and Honesty Outcome (Truth vs. Lie) among Communicators who helped others (Prosocial Outcome: Help) in the Prosociality Context. There was a significant main effect of Honesty Outcome ($F(1,195)$ = 4.26, p = .040) such that Communicators who ultimately ended up telling the truth (M = 4.99, SD = 1.21) were seen as more moral than those who lied (M = 4.64, SD = 1.22). Unlike the Honesty Context, however, we found no significant main effect of Decision Strategy ($F(1,195)$ = 0.23, p = .633), nor did we find a significant Decision Strategy x Honesty Outcome interaction ($F(1,195)$ = 0.10, p = .753). In other words, there does not seem to be a preference for unconditional prosociality.

- ***Other Measures***

Following our preregistration, we ran a 2x2x2x2 ANOVA on judgments of trust, benevolence, integrity, and honesty using Context, Decision Strategy, Honesty Outcome, and Prosocial Outcome as factors. The descriptive statistics are available in Table S22. We report the full 2x2x2x2 ANOVA results in Table S23.

We also ran a logistic regression using passing behavior as our dichotomous DV (1 = pass, 0 = keep) using context (1 = Honesty, 0 = Prosociality), decision strategy (1 = Looking, 0 = Unconditional Behavior), honesty outcome (1 = Truth, 0 = Lie), and prosocial outcome (1 = Help, 0 = Harm), and all their interactions as categorical IV's. The regression results on trust behavior (pass in the Trust Game) are displayed in Table S24.

Table S22.

Descriptive Results of Non-Morality Measures from Study S7

	Honesty Context							
	Unconditional behavior				Looking			
	Truth		Lie		Truth		Lie	
Dependent Variables	Help	Harm	Help	Harm	Help	Harm	Help	Harm
Trustworthiness	4.27 (1.41)	4.21 (1.50)	3.36 (1.54)	2.98 (1.52)	4.75 (1.32)	3.22 (1.73)	4.33 (1.41)	2.35 (1.57)
Benevolence	4.83 (0.98)	4.57 (1.18)	3.89 (1.45)	3.51 (1.35)	4. 98 (1.07)	3.42 (1.67)	5.16 (1.08)	2.60 (1.59)
Integrity	4.74 (0.94)	4.62 (0.97)	3.93 (1.23)	3.69 (1.20)	4.73 (0.97)	4.13 (1.43)	4.61 (1.05)	3.39 (1.40)
Honest	5.16 (0.94)	5.02 (1.02)	3.55 (1.77)	3.38 (1.60)	5.17 (1.03)	4.23 (1.77)	4.38 (1.52)	2.73 (1.71)

	Prosociality Context							
	Unconditional behavior				Looking			
	Truth		Lie		Truth		Lie	
Dependent Variables	Help	Harm	Help	Harm	Help	Harm	Help	Harm
Trustworthiness	4.39 (1.44)	2.61 (1.71)	3.97 (1.37)	2.52 (1.53)	4.53 (1.52)	3.64 (1.46)	3.97 (1.37)	2.36 (1.53)
Benevolence	4.83 (1.46)	2.71 (1.62)	4.75 (1.13)	2.79 (1.55)	4.84 (1.15)	3.81 (1.53)	4.91 (1.30)	2.86 (1.69)
Integrity	4.88 (1.00)	3.53 (1.30)	4.40 (1.08)	3.52 (1.26)	4.95 (1.26)	4.81 (1.01)	4.64 (1.21)	3.42 (1.55)
Honest	4.69 (1.48)	3.17 (1.88)	4.50 (1.29)	3.11 (1.80)	5.31 (1.22)	5.20 (1.50)	4.48 (1.60)	2.83 (1.74)

Table S23.

Full ANOVA Results from Study S7

	Moral		Trust		Integrity		Benevolence		Honest	
	F	p	F	p	F	p	F	p	F	p
Context	1.02	.314	2.05	.153	0.19	.662	3.38	.066	0.15	.697
Decision Strategy	1.18	.279	1.62	.203	**4.13**	**.043**	.826	.364	**4.07**	**.044**
Honesty Outcome	**59.48**	**<.001**	**42.08**	**<.001**	**49.70**	**<.001**	**20.16**	**<.001**	**108.45**	**<.001**
Prosocial Outcome	**152.29**	**<.001**	**134.97**	**<.001**	**72.20**	**<.001**	**231.49**	**<.001**	**77.08**	**<.001**
Context x Decision Strategy	**6.99**	**.008**	2.78	.096	**5.51**	**.019**	**6.25**	**.013**	**11.80**	**.001**
Context x Honesty Outcome	1.39	.239	2.36	.125	0.36	.551	**5.05**	**.025**	**5.84**	**.016**
Context x Prosocial Outcome	**10.58**	**.001**	**5.56**	**.019**	**4.35**	**.037**	**9.48**	**.002**	**4.19**	**.041**
Decision Strategy x Honesty Outcome	0.45	.500	0.09	.763	0.23	.634	0.34	.562	**5.29**	**.022**
Decision Strategy x Prosocial Outcome	**5.92**	**.015**	**9.27**	**.002**	0.78	.376	**10.10**	**.002**	1.73	.188
Honesty Outcome x Prosocial Outcome	**7.05**	**.008**	2.75	.098	3.99	.046	**6.32**	**.012**	**6.24**	**.013**
Context x Decision Strategy x Honesty Outcome	**10.50**	**.001**	**5.21**	**.023**	**9.51**	**.002**	**8.18**	**.004**	**20.56**	**<.001**
Context x Decision Strategy x Prosocial Outcome	**27.93**	**<.001**	**17.42**	**<.001**	**11.68**	**.001**	**32.45**	**<.001**	**15.76**	**<.001**
Context x Honesty Outcome x Prosocial Outcome	0.009	.924	0.03	.855	0.02	.881	0.12	.731	0.62	.432
Decision Strategy x Honesty Outcome x Prosocial Outcome	**9.83**	**.002**	2.82	.094	**9.19**	**.003**	**7.077**	**.008**	**7.38**	**.007**

Context x Decision Strategy x Honesty Outcome x Prosocial Outcome	1.84	.175	1.86	.173	2.28	.131	1.58	.692	1.31	.254

Table S24.

Regression Results Predicting Trust Behavior (Pass in the Trust Game) in Study S7

	b	*SE*	*p*
Constant	-0.81	0.30	.007
Context	-0.08	0.42	.850
Decision Strategy	-0.05	0.44	.915
Honesty Outcome	0.41	0.43	.343
Prosocial Outcome	1.12	0.41	.006
Context by Decision Strategy	0.27	0.61	.652
Context by Honesty Outcome	-0.10	0.59	.871
Context by Prosocial Outcome	0.62	0.59	.299
Decision Strategy by Honesty Outcome	-0.05	0.60	.936
Decision Strategy by Prosocial Outcome	-0.04	0.59	.945
Honesty Outcome by Prosocial Outcome	-0.02	0.60	.970
Context by Decision Strategy by Honesty Outcome	0.58	0.84	.494
Context by Decision Strategy by Prosocial Outcome	-1.49	0.84	.076
Context by Honesty Outcome by Prosocial Outcome	-0.38	0.85	.650
Decision Strategy by Honesty Outcome by Prosocial Outcome	0.23	0.85	.785
Context by Decision Strategy by Honesty Outcome by Prosocial Outcome	-0.25	1.19	.835

Notes. The displayed results are from a logistic regression using passing behavior as our dichotomous DV (1 = pass, 0 = keep) and context (1 = Honesty, 0 = Prosociality), decision strategy (1 = Looking, 0 = Unconditional Behavior), honesty outcome (1 = truth, 0 = lie), prosocial outcome (1 = help, 0 = harm), and all of their interactions as categorical independent variables.

Figure S2

Moral Judgment in the Honesty and Prosociality Contexts (Study S7)
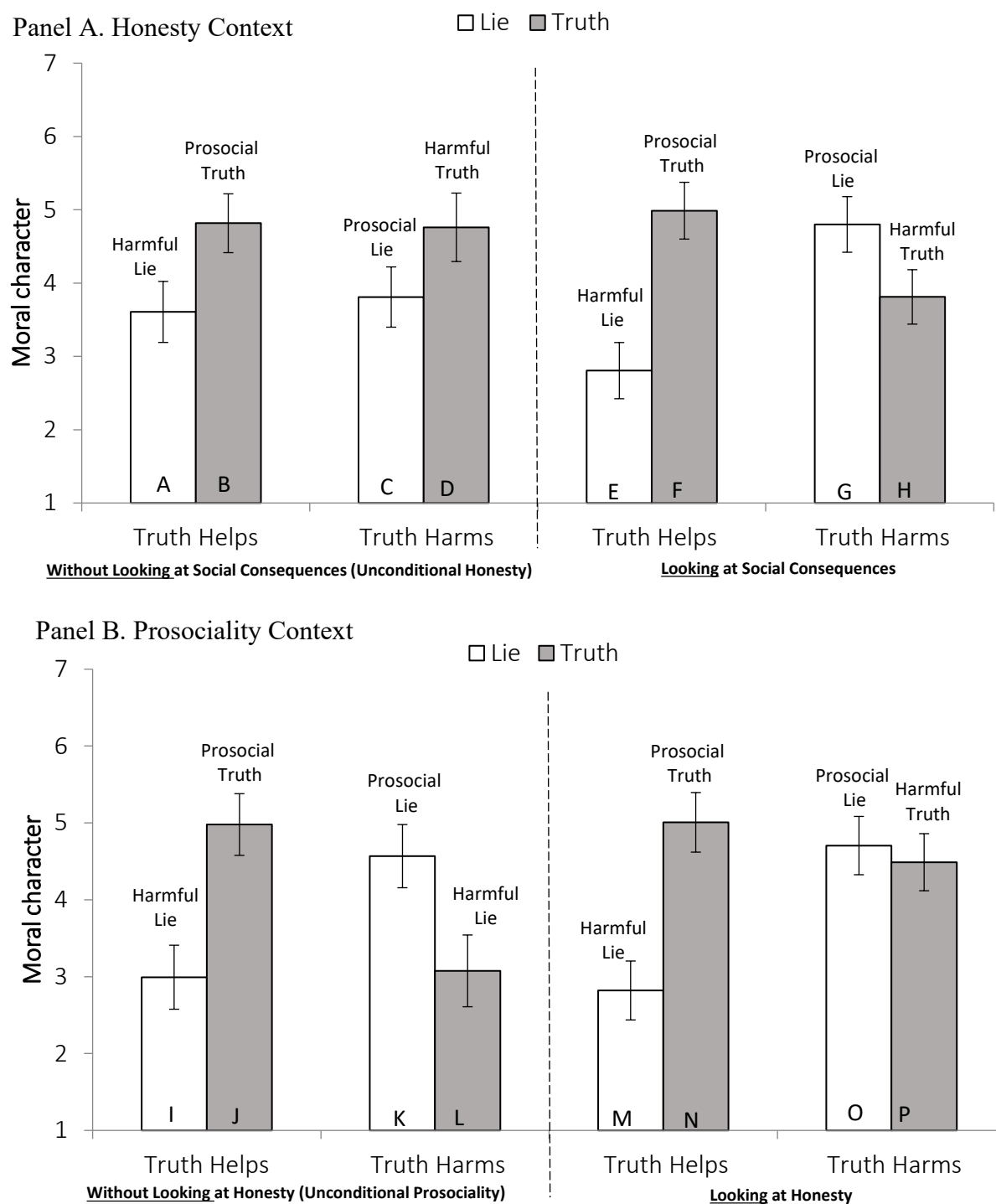
Notes. Study S7 results. Error bars represent 95% confidence intervals around the mean.

Discussion

Study S7 makes several contributions. First, it replicates the results of Study 5 of the main manuscript: Unconditional honesty is not seen as more ethical than Looking if it is certain that Looking will not lead to harmful lies. Study S7, additionally, offers new insights into the nature of these effects. Examining the full pattern of results in the Honesty context sheds light on how honest and deceptive behaviors are judged as a function of the information available at the moment of decision-making (i.e., certainty). In the Looking condition of the Honesty context, there is a significant Honesty Outcome x Prosocial Outcome interaction ($F(1, 207) = 4.572$, $p = .034$). Prosocial truth-telling (Column F in Figure S2, Panel A) is seen as significantly more ethical than harmful lying (Column E in Figure S2 Panel A, $t(106) = 8.25$, $p < .001$), and prosocial lying (Column G in Figure S2, Panel A) is seen as more ethical than harmful truth-telling (Column H in Figure S2, Panel A, $t(101) = 3.57$, $p = .001$). In other words, consistent with Studies 1 and S1 and prior work, the moral judgment of deception depends on its social consequences, if decision makers are certain of these consequences at the time of making the decision. Furthermore, consistent with our account, these judgments are asymmetric such that harmful lying is clearly the worst outcome.

However, in the Unconditional Behavior condition, when communicators are uncertain about the consequences of honesty, there is not a significant Lie x Outcome interaction ($p = .660$). There is just a main effect of Honesty ($F(1,190) = 41.68$, $p < .001$). If one simply follows the moral rule of honesty, it does not matter if that rule leads to (foreseeable) harm. Whereas the studies in the main manuscript demonstrate that honesty is rewarded a priori (before harm is realized), Study S7 finds that it is also rewarded post hoc, after it causes harm. This finding suggests, consistent with Study S1, that preferences for unconditional honesty are not driven by a failure to anticipate the harm it may cause. Intentional ignorance of honesty's consequences seems to be rewarded. We replicate these patterns of results in Study S9 (see SOM 2.9); albeit some comparisons vary in significance.

Study S7 also provides further insight into the generalizability of preferences for unconditional moral rule-following. As in Study S6, we do not find that people have a preference for unconditional prosociality relative to Looking, when the truth is uncertain. Nonetheless, it is important to note that unconditional prosociality is *no worse* than Looking. This suggests that unconditional moral rule-following may still be valued as a general strategy across contexts.

2.8. Study S8: Rely-or Verify

In several of our supplemental studies, we used The Choice Game, a standard trust game, to explore whether observers would trust communicators who engaged in Unconditional Honesty more than they trusted those who engaged in Looking. However, previous research has shown that the standard trust game primarily captures benevolence-based trust. In the present study, we use the Rely-or-Verify (RV) Game (adapted from Levine & Schweitzer, 2015) to test whether unconditional honesty increases trust when the trust measure is sensitive to judgments of integrity. Given that unconditional honesty does not reflect a concern for benevolence per se, we expected that unconditional honesty might increase trust in the RV Game even if it did not in the Choice Game (Trust Game).

2.8.1. Method for Study S8

Participants
- $N = 434$ ($M_{age} = 34.62$; $SD_{age} = 10.51$; 242 males, 188 females, 4 prefer to self-describe)
- Collected on Prolific

Procedure
- Participants were assigned to a partner and were told they would play a game with that partner. In the results section, we will refer to this partner as the Communicator.
- Participants learned about the Rely-or-Verify Game. In the game, they were assigned to the role of BLUE Player and their partner was assigned to the role of RED Player. Participants were told that they had to report whether the amount of money in a jar was odd or even and that their report would affect theirs and their partner's bonus payments.
  - Participants were told that RED Player knew the amount of money in the jar and would send a message that the amount was either odd or even.
  - Then, participants would have to RELY or VERIFY. Based on their decision, a series of outcomes were possible.
    1. If **RED Player** gave you accurate information and you **RELY** on it, you will correctly report whether the amount of money in the jar is ODD or EVEN, and you will earn $1.50. **RED Player** will earn $0.75.
    2. If **RED Player** gave you inaccurate information and you **RELY** on it, you will incorrectly report whether the amount of money in the jar is ODD or EVEN, and you will earn $0. **RED Player** will earn $1.50.
    3. If **RED Player** gave you accurate information and you **VERIFY** it, you will earn $1 ($1.50 for the correct answer - $0.50 cost of verification) and **RED Player** will earn $0.50.
    4. If **RED Player** gave you *in*accurate information and you **VERIFY** it, you will earn $1 ($1.50 for the correct answer - $0.50 cost of verification) and **RED Player** will earn $0.
- After learning about the Rely-or-Verify Game, participants then learned about the Coin Flip Game (see Study 2) and learned how their partner behaved in the Coin Flip Game. Specifically, they were assigned to a condition in a 2 (Decision Strategy: Unconditional Honesty vs. Looking) between-subjects design.
- Participants played the Rely-or-Verify Game with their partner and completed attitudinal DV measures.
- Lastly, participants answered demographic questions.

Dependent Variables

- Our main dependent variable was behavioral trust as measured by the propensity for participants to rely on their partner's report in the Rely-or-Verify Game.
- Our other focal dependent variable was ratings of the Communicator's morality (α = .962) using the same items as previous studies.
- We also measured judgments of the Communicator's trustworthiness using a 3-item composite of "I trust [Communicator]," "I am willing to make myself vulnerable to [Communicator]," and "I am confident that [Communicator] sent me an accurate message" (α = .894).

Analyses
- We conducted a logistic regression on trust (1 = rely, 0 = verify) using Decision Strategy (1 = Unconditional Honesty, 0 = Looking) as an independent variable, and followed this analysis with a chi-squared test examining whether integrity-based trust is higher depending on the Communicator's Decision Strategy.
- We also conducted a series of independent samples t-tests examining judgments of morality and trustworthiness between each Decision Strategy.

2.8.2. Results & Discussion for Study S8

Results
Rely-or-Verify Game

The logistic regression revealed that Decision Strategy did not predict the extent to which participants chose to rely on the Communicators message in the Rely-or-Verify Game (p = .388). The follow-up chi-squared test indicated participants were equally likely to rely on messages from Communicators who engaged in Unconditional Honesty (51.6%) as Communicators who engaged in Looking (47.5%; χ^2 = 0.75, p = .388).

Morality

Participants reported that Communicators who engaged in Unconditional Honesty (M = 4.98, SD = 1.07) were more moral than Communicators who engaged in Looking (M = 4.66, SD = 1.30; t(432) = 2.82, p = .005).

Other Measures

Unconditional Honesty Communicators (M = 4.18, SD = 1.55) were also judged to be more trustworthy than Looking Communicators (M = 3.84, SD = 1.65; t(432) = 2.21, p = .028).

Discussion

In this study, we replicated our past work that communicators who engage in unconditional honesty are judged as more moral and trustworthy than communicators who engage in looking at the consequences of the truth. However, we did not find evidence that unconditional honesty increases integrity-based trust.

2.9. Study S9: Unconditional Honesty 2x2x2

Study S9 was a replication of Study S7 without the Prosociality Context, examining the judgment of Unconditional Honesty vs. Looking once the outcomes of those decisions were known.

2.9.1. Method for Study S9

Participants
- N = 804 (M_{age} = 41.59; SD_{age} = 12.87; 430 males, 366 females, 2 prefer to self-describe, 6 prefer to not disclose)
- Collected on MTurk

Procedure

- Participants learned about a Communicator who had previously participated in a study in which they played "the Number Game" (see Table S25).
- We informed participants of the ultimate choice the Communicator made, including the Communicator's decision strategy (i.e., Unconditional Behavior vs. Looking), the Honesty Outcome (whether the Communicator ultimately told the truth or lied), and the Prosocial Outcome (whether the Communicator ultimately helped or harmed another person). That is, participants in this study were randomly assigned to a condition from a 2 (Decision Strategy: Unconditional Behavior vs. Looking) x 2 (Honesty Outcome: Truth vs. Lie) x 2 (Prosocial Outcome: Help vs. Harm) between-subjects design.

Table S25.

Experimental stimuli from Study S9 (Number Game Instructions)

In this study, you will learn about a decision a previous participant made in a game called the "**NUMBER GAME**".

To make sure that you understand the **NUMBER GAME**, we will ask you to read the instructions and answer comprehension questions.

Please read the instructions carefully.
We will now pair you with another MTurk participant who previously played the **NUMBER GAME**.

(This could take a few seconds.)

You have been assigned to Person [X].

Neither of you knows the identity of the other.
Press "Next" to learn the **NUMBER GAME.**

Next, you will learn about what your partner (Person [X]) did in a previous game called the **NUMBER GAME**.
Please wait while we load your partner's (Person [X]'s) instructions and choice from the **NUMBER GAME**...

(This could take a few seconds.)

In the **NUMBER GAME,** your partner (Person [X]) was in the role of **DECIDER**.

Your partner was paired with a different person (not you), who was in the role of **RECEIVER**. This RECEIVER played a previous game and earned $1. However, this RECEIVER also knew that their $1 earning might be affected by another person's (your partner (Person [X]'s) decision.

On the next page are the exact instructions that your partner saw as a **DECIDER**.

Below are the instructions your partner (Person [X]) saw in the NUMBER GAME:

NUMER GAME – DECIDER INSTRUCTIONS
You will get to choose an allocation for your partner, the RECIEVER. The allocation will depend on your choice and the outcome of a random number generator.

First, you will generate a number between 1 and 9. Inclusive. The number will either be: 1, 2, 3, 4, 5, 6, 7, 8, or 9.

Then, you will have to report whether the number is ODD or EVEN.

If you report the number is ODD, your partner will be paid according to Option A.
If you report the number is EVEN, your partner will be paid according to Option B.
- Option A) $X for the RECEIVER
- Option B) $Y for the RECEIVER

You will not know the values of X and Y when you make your choice.
We will now learn how your partner **(Person [X])** responded in the **NUMBER GAME**.
Here is what they saw when they used the random number generator:

[Presented in Looking / Truth / Help Condition; Looking / Lie / Hurt Condition; Not Look / Truth / Help Condition; Not Look Lie / Hurt Condition]
A random number has been generated.
The number is: 7.
Please remember this number. Feel free to write it down.

[Presented in Look / Truth / Hurt Condition; Look / Lie / Help Condition; Not Look / Truth / Hurt Condition; Not Look / Lie / Help Condition]
A random number has been generated.
The number is: 4.
Please remember this number. Feel free to write it down.

Your partner (Person [X]) then saw the following instructions:

You will now make a decision.

You can report **ODD**, **EVEN**, or **You can choose to find out the values of $X and $Y before making a decision.**

If you choose to find out what $X and $Y are before making a decision, you will then be directed to a screen that informs you of the values of $X and $Y. Then, you will have to report either:

- The random number generator generated an ODD number, which will result in $X, or
- The random number generator generated an EVEN number, which will result in $Y.

We will now learn how your partner (Person [X]) responded in the **NUMBER GAME**.

Seeing that the random number generated produced an [ODD / EVEN] number, here is how your partner (Person [X]) responded.

[Presented in Not Look / Truth / Help Condition; Not Look / Lie / Help Condition]

Please make a decision. This will affect your partner's bonus payment.

The random number generator generated an **ODD** number.
This decision will lead to allocation Option A) $X for the RECEIVER

The random number generator generated an **EVEN** number.
This decision will lead to allocation Option B) $Y for the RECEIVER

I choose not to report a number
I would like to find out the values of X and Y and the make a decision

[Presented in Not Look / Truth / Hurt Condition; Not Look / Lie / Hurt Condition]

Please make a decision. This will affect your partner's bonus payment.

The random number generator generated an **ODD** number.
This decision will lead to allocation Option A) $X for the RECEIVER

The random number generator generated an **EVEN** number.
This decision will lead to allocation Option B) $Y for the RECEIVER

I choose not to report a number
I would like to find out the values of X and Y and the make a decision

[Presented in Looking / Truth / Help Condition; Looking / Truth / Hurt Condition; Looking / Lie / Help Condition; Looking / Lie / Hurt Condition]

Please make a decision. This will affect your partner's bonus payment.

The random number generator generated an **ODD** number.
This decision will lead to allocation Option A) $X for the RECEIVER

The random number generator generated an **EVEN** number.
This decision will lead to allocation Option B) $Y for the RECEIVER

I choose not to report a number
I would like to find out the values of X and Y and the make a decision

Dependent Variables

- The main dependent variable was morality of the Communicator; we used the same three-item composite as previous studies ($\alpha = .972$).

Analyses

- To examine whether Unconditional Honesty is seen as more moral than Looking, holding outcomes constant, we conducted a two-way ANOVA examining the effect of Decision Strategy (Looking vs. Unconditional Behavior) and Prosocial outcome (Help vs. Harm) on moral judgments for Communicators who told the truth (Honesty Outcome: Truth).
- To examine whether Unconditional Honesty is more moral than Looking when looking results in prosocial outcomes, we conducted a t-test examining the moral judgment of Unconditional Honesty (Decision Strategy: Unconditional Behavior, Honesty Outcome: Truth) to the moral judgment of Prosocial Looking (Decision Strategy: Looking, Prosocial Outcome: Help).
- For completeness, we also conducted a full 2 (Decision Strategy: Unconditional Behavior vs. Looking) x 2 (Honesty Outcome: Truth vs. Lie) x 2 (Prosocial Outcome: Help vs. Harm) ANOVA (see Table S27).

2.9.2. Results & Discussion for Study S9

Results
Table S26 depicts the means and standard deviations of all cells. Table S27 depicts ANOVA results.

Table S26.

Descriptive Statistics for Study S9

	Unconditional Behavior				**Looking**			
	Lie Harm	**Truth Help**	**Lie Help**	**Truth Harm**	**Lie Harm**	**Truth Help**	**Lie Help**	**Truth Harm**
Morality	2.70 (1.55)	5.24 (1.30)	3.64 (1.69)	4.78 (1.61)	1.90 (1.15)	5.56 (1.28)	4.37 (1.68)	4.06 (1.98)

Is Unconditional Honesty seen as more moral than Looking, holding outcomes constant?

We conducted a two-way ANOVA examining the effect of Decision Strategy (Looking vs. Unconditional Behavior) and Prosocial outcome (Help vs. Harm) on moral judgments for Communicators who told the truth (Honesty Outcome: Truth). We did not find a main effect of Decision Strategy (p = .205). However, we did find a main effect of Prosocial outcome ($F(1,395)$ = 38.86, p < .001) such that Communicators who helped (M = 5.40, SD = 1.30) were viewed as more moral than those who harmed (M = 4.42, SD = 1.84).

We also found an interaction between Decision Strategy and Prosocial Outcome ($F(1,395)$ = 10.88, p = .001) such that Communicators who Looked and helped (M = 5.56, SD = 1.28) were rated similarly to those who engaged in Unconditional Honesty and helped (M = 5.24, SD = 1.30, p = .083). Communicators who Looked and caused harm were rated as less moral (M = 4.06, SD = 1.98) than those who engaged in Unconditional Honesty and caused harm (M = 4.78, SD = 1.61, p = .005).

Is Unconditional Honesty seen as more moral than Looking, when Looking results in prosocial outcomes?

We conducted a t-test examining the moral judgment of Unconditional Honesty (Decision Strategy: Unconditional Behavior, Honesty Outcome: Truth) to the moral judgment of Prosocial Looking (Decision Strategy: Looking, Prosocial Outcome: Help). This test revealed that when prosocial outcomes were realized, Unconditional Honesty was judged as moral as Looking ($M_{Unconditional\ Honesty}$ = 5.01, $SD_{Unconditional\ Honesty}$ = 1.48 vs. $M_{Looking}$ = 4.94, $SD_{Looking}$ = 1.6, p = .645), consistent with Study 5 of the main manuscript.

Table S27.

Full ANOVA Results on Moral Judgments in Study S9

	F	*p*
Decision Strategy	1.19	.276
Honesty Outcome	**257.68**	**< .001**
Prosocial Outcome	**150.41**	**< .001**
Decision Strategy x Prosocial Outcome	**34.09**	**< .001**
Decision Strategy x Honesty Outcome	0.54	.464
Honesty Outcome x Prosocial Outcome	**11.00**	**< .001**
Decision Strategy x Honesty Outcome x Prosocial Outcome	1.22	.270

Discussion

In this study, we find support that Looking is viewed as similarly moral as Unconditional Honesty if it is certain that Looking will not lead to harmful lies, consistent with Study 5 of the manuscript, as well as Study S7.

2.10. Study S10: Unconditional Honesty, Looking, and Prosocial Intentions

2.10.1. Method for Study S10

Participants
- $N = 598$ ($M_{age} = 35.17$, $SD_{age} = 11.61$; 298 males, 300 females)
- Collected on MTurk

Procedure
- Participants learned about a Communicator who played an adapted version of the Coin Flip Game from Study 2.
- However, we made a key change to the game. We informed participants that Communicators had the opportunity to provide an explanation for their communication strategy (Unconditional Honesty or Looking). Participants saw a free-response text box below the Communicator's choice. In the Unconditional Honesty and Looking conditions, this box was left blank. In the Looking-With-Intention condition, however, the Communicator wrote "If A is $1, I'll report heads. If A is -$1, I'll report tails." This statement clarified that the Communicator was prosocially motivated; the Communicator chose Looking because they intended to report the coin flip outcome that would lead their partner to gain money (regardless of whether or not the report would be honest).
- Participants were randomly assigned to one of three possible conditions (Decision Strategy: Unconditional Honesty vs. Looking vs. Looking-With-Prosocial Intention).

Dependent Variables
- Our primary dependent variable was judgments of the Communicator's morality using the same three-item composite as prior studies ($\alpha = .959$).
- Participants also reported whether they would like to play the Message Game in a future study with the Communicator or a novel partner (Communicator vs. Different Partner), and they also rated the extent to which they would like to play the Message Game with the Communicator (1 = Not at all, 7 = Very much). The Message Game was described as a study that "could involve having intimate conversations with your partner, completing tasks with your partner, and dividing resources with them." We conceptualized activities in the Message Game as different trusting behaviors.
- Finally, participants answered a series of questions about the target Communicator and the target Communicator's decision (1 = Strongly disagree, 7 = Strongly agree):
  - To what extent do you think your partner would:
    - tell lies to benefit themselves
    - tell lies that intentionally harm others
    - tell lies in arbitrary circumstances
    - tell well-meaning lies that actually cause harm
  - How likely is it that your partner will:
    - tell a lie in the Coin Flip Game that leads their partner to lose $1 (*antisocial lie*)
    - tell a lie in the Coin Flip Game that leads their partner to gain $1 (*prosocial lie*)
    - a truth in the Coin Flip Game that leads their partner to lose $1 (*antisocial truth*)
    - a truth in the Coin Flip Game that leads their partner to gain $1 (*prosocial truth*)
  - How ethical would it be for your partner to:
    - tell a lie in the Coin Flip Game that leads their partner to lose $1
    - tell a lie in the Coin Flip Game that leads their partner to gain $1
    - tell a truth in the Coin Flip Game that leads their partner to lose $1
    - tell a truth in the Coin Flip Game that leads their partner to gain $1

Analyses
- We conducted a one-way ANOVA on judgments of morality, the extent to which participants want to play a future game with their partner, and the 12 items about their partner and their partner's decision, using Decision Strategy as a factor.
- We also conducted a chi-square analysis to measure the extent to which individuals prefer to work on another study with their partner or a novel partner based on condition.

2.10.2. Results & Discussion for Study S10

Morality

We found a main effect of Decision Strategy on judgments of morality ($F(2,595) = 23.76$, $p <$ .001). Participants judged Communicators in the Unconditional Honesty condition ($M = 5.47$, $SD = 1.06$) as more ethical than Communicators in the Looking condition ($M = 4.89$, $SD = 0.99$; $p < .001$) and Looking-With-Prosocial Intention condition ($M = 4.86$, $SD = 0.94$; $p < .001$). Participants judged Communicators in the Looking and Looking-With-Prosocial Intention conditions as equally moral ($p =$ .777). In other words, clarifying one's prosocial intent for Looking did not attenuate the social benefits of Unconditional Honesty.

Other Measures

We found an effect of Decision Strategy on the extent to which participants wanted to play another game with the Communicator ($F(2,595) = 10.62$, $p < .001$). Post-hoc tests revealed that participants in the Unconditional Honesty condition had a stronger preference to play another game with the Communicator than those in the Looking condition ($p < .001$) or Looking-With-Prosocial Intention condition ($p < .001$). Participants had similar preferences to continue working with the Communicator in the Looking and Looking-With-Prosocial Intention ($p = .418$). However, the effect of condition on participants' dichotomous choice to play the Message Game with the Communicator versus another study participant was not significant at the $p < .05$ threshold, though it trended in the same direction ($\chi^2 = 5.34$, $p = .069$).

We found an effect of Decision Strategy on the extent to which participants believed Communicators would tell an antisocial lie (variable: LieToLose) ($F(2,595) = 15.22$, $p < .001$), prosocial lie (variable: LieToGain) ($F(2,595) = 6.52$, $p = .002$), antisocial truth (variable: TruthToLose) ($F(2,595) =$ 6.71, $p = .001$), and prosocial truth (variable: TruthToGain) ($F(2,595) = 11.97$, $p < .001$) in the Coin Flip Game. Post-hoc tests revealed that participants in the Unconditional Honesty condition believed the Communicator was less likely to tell antisocial lies and prosocial lies but more likely to tell prosocial truths and antisocial truths than those in the Looking or Looking-With-Prosocial Intention conditions (*ps* < .003). Participants in the Looking and Looking-With-Prosocial Intention conditions expected the Communicator to be equally likely to tell an antisocial lie, prosocial lie, antisocial truth, and prosocial truth *(ps ≥ .749)*. These results suggest that our "Looking with Prosocial Intention" manipulation did not change beliefs about the Communicator's behavior.

We did not find an effect of Decision Strategy on the extent to which participants judged the ethicality of telling an antisocial lie, prosocial lie, antisocial truth, or prosocial truth in the Coin Flip Game (*p*s > .448).

We found an effect of Decision Strategy on the extent to which participants believed the Communicator would lie to benefit themselves ($F(2,595) = 19.73$, $p < .001$), lie to harm others $F(2,595) =$ 7.92, $p < .001$), lie in arbitrary situations $F(2,595) = 9.16$, $p < .001$), and tell well-meaning lies $F(2,595) =$ 10.64, $p < .001$). Participants in the Unconditional Honesty condition believed the Communicator would be less likely to tell lies to benefit themselves, lie to harm others, lie in arbitrary situations, and tell well-meaning lies than those in the Looking and Looking-With-Prosocial Intention conditions (*p*s < .058). Communicators in the Looking condition were judged as equally likely as those in the Looking-With-Intention condition to lie to benefit themselves, lie to harm others, and lie in arbitrary situations (*p*s >

.061). However, participants in the Looking-With-Prosocial Intention (vs. Looking) condition believed that the Communicator was *less* likely to tell well-meaning lies (*p* = .008).

Descriptive statistics for all measures in Study S10 are reported in Table S28.

Table S28.

Descriptive Statistics in Study S10

Dependent variable	Unconditional Honesty	Looking	Looking-With-Intention
Want to Play with Target in Future Study	5.13 (1.46)	4.53 (1.39)	4.64 (1.33)
Lie to benefit self	3.04 (1.48)	3.89 (1.34)	3.63 (1.31)
Lie to intentionally harm others	2.64 (1.40)	3.14 (1.31)	3.07 (1.34)
Lie in arbitrary situations	3.05 (1.51)	3.63 (1.35)	3.45 (1.35)
Tell well-meaning lie	3.03 (1.47)	3.64 (1.28)	3.28 (1.30)
Tell a lie that leads partner to lose $1 (*antisocial lie*)	2.75 (1.52)	3.43 (1.42)	3.48 (1.46)
Tell a lie that leads partner to gain $1 (*prosocial lie*)	3.29 (1.64)	3.79 (1.52)	3.77 (1.49)
Tell a truth that leads partner to lose $1 (*antisocial truth*)	5.12 (1.50)	4.64 (1.42)	4.69 (1.42)
Tell a truth that leads partner to gain $1 (*prosocial truth)*	5.61 (1.34)	5.03 (1.37)	5.06 (1.29)
Ethical to antisocial lie	2.35 (1.48)	2.41 (1.45)	2.50 (1.47)
Ethical to prosocial lie	3.44 (1.73)	3.53 (1.76)	3.57 (1.83)
Ethical to antisocial truth	5.26 (1.60)	5.34 (1.57)	5.36 (1.55)
Ethical to prosocial truth	5.88 (1.32)	5.74 (1.36)	5.88 (1.27)
Percent Choosing Same Communicator for Message Game	81.6%	71.8%	77.4%

Discussion

The results of Study S10 suggest that stating one's prosocial intentions when Looking is not sufficient to attenuate observers' moral preference for unconditional honesty. However, it is possible that participants in our study did not fully understand our manipulation of Prosocial Intention, or already assumed some degree of prosocial intention in the Looking condition. Participants did not rate the communicator as more likely to tell the prosocial lie *or* prosocial truth in the Looking-with-Prosocial intention condition, relative to the Looking condition, and they rated the communicator as less likely to tell well-meaning lies in the Looking-with-Prosocial intention condition, relative to the Looking condition. We address these concerns in Study 5 of the main manuscript.

3. Summary of Communicators' Decisions

Table S29.

Communicator Decision Strategies in Studies S3a, S4a, and S6a

	Communicator's Initial Decision Strategy		
	Unconditional Honesty	Unconditional Lying	Looking
When Honesty has Personal Consequences (Study S3a)	49.1%	113%	39.6%
When Honesty has Social Consequences (Study S3a)	60.4%	7.5%	32.1%
When Honesty has Both Personal and Social Consequences (Study S3a)	54.7%	1.9%	43.4%
When Harm Associated with Honesty is Salient (Study S4a)	51.9%	0.9%	47.2%
When Harm Associated with Honesty is Not Salient (Study S4a)	64.9%	0.0%	35.1%
General Honesty Context (Study S6a)	62.0%	11.0%	27.0%

Three of our supplemental studies provide initial insight into the frequency with which communicators engage in unconditional honesty (see SOM 2.3, 2.4, and 2.6). Across these studies, the modal communicator behavior was unconditional honesty. However, the preference for unconditional honesty weaker when there were only personal, rather than social, consequencs for honesty and when the harm of honesty was less salient.

We also examine Communicators' decision strategies in Study 3 (see SOM 1.3.2).

4. Preregistration Information

4.1. Main Studies

Study 1: Moral judgment of Unconditional Honesty
- Not preregistered

Study 2: The reputational consequences of unconditional honesty
- Link: https://aspredicted.org/blind.php?x=z28qq7

Study 3: Harmful lies and the preference for Unconditional Honesty
- Link: https://aspredicted.org/48S_9DQ

Study 4: The moral preference for Unconditional Honesty across relationships
- Link: https://aspredicted.org/YDK_TX2

Study 5: Judgments of Unconditional Honesty After Eliminating the Possibility of Error
- Link: https://aspredicted.org/blind.php?x=52vw62

4.2. Supplemental Studies

Study S1: Moral judgment of Unconditional Honesty (Within Subjects)
- Link: https://aspredicted.org/blind.php?x=3f3g5s

Study S2: Judgments of Unconditional Honesty when Honesty has Personal vs. Social Consequences
- Link: http://aspredicted.org/blind.php?x=yg2ay3

Study S3: Judgments of Unconditional Honesty when Honesty has Personal *and* Social Consequences
- Link: https://aspredicted.org/blind.php?x=en2r8v

Study S4: Judgments of Unconditional Honesty and Harm Salience
- Not preregistered

Study S5: Judgments of Unconditional Honesty and Perceptions of Selfishness
- Link: https://aspredicted.org/blind.php?x=2337um

Study S6: Unconditional Prosociality
- Link: https://aspredicted.org/blind.php?x=ks9th8

Study S7: Unconditional Prosociality 2
- Link: https://aspredicted.org/blind.php?x=9sq63e

Study S8: Rely-or-Verify
- Link https://aspredicted.org/blind.php?x=vz2ds8

Study S9: Unconditional Honesty 2x2x2
- Link: https://aspredicted.org/blind.php?x=iu38c6

Study S10: Unconditional Honesty, Looking, and Prosocial Intentions
- Link: https://aspredicted.org/blind.php?x=bv32g7