**On being honest about dishonesty:**

**The social costs of taking nuanced (but realistic) moral stances**

**SI Appendix**

**Table of Contents**

# 1. Additional Results for Studies 1-5

## 1.1 Study 1 Additional Results

We ran a series of 2 (proclamation) x 2 (frame) x 3 (perspective) ANOVAs on each of our dependent variables in Study 1: voting intentions, morality, future honesty, and hypocrisy. We discuss the significant main effect of proclamation on each variable in the main manuscript, which are the results of primary theoretical interest. No other effects, including interactions effects, were significant at the $p < .05$ threshold. However, we include descriptive statistics for all dependent variables in Table SI-1 on page 4, and we discuss additional effects of theoretical interest below.

**Additional Attitudinal Results**

The two-way and three-way interactions between proclamation, frame, and perspective were not significant at the $p < .05$ threshold in any of our ANOVA models. Nevertheless, examining trends in the three-way proclamation x frame x perspective interaction can still be informative for future work. Specifically, exploring the three-way interaction in the models predicting moral evaluations ($F(2, 1009) = 1.74$, $p = .176$, $\eta_p^2 < .01$) and voting intentions ($F(2, 1009) = 2.33$, $p = .097$, $\eta_p^2 < .01$) is useful for understanding when flexible proclamations may not undermine judgments of one's character.

When breaking down these interactions, we find that absolute honesty proclamations were beneficial for voting intentions and moral evaluations, relative to flexible proclamations, if the proclamation was made from the second-person or third-person perspective (with any framing) or if the proclamation was made from the first-person perspective with a truth frame. However, when the communicator used the first-person perspective with the lie frame ("*As a politician, I would never lie"*), the effect of proclamation on voting intentions ($t(172) = -1.06$, $p$

= .291, $d$ = -0.16) and moral evaluations ($t$(172) = -1.08, $p$ = .282, $d$ = -0.16) was no longer significant, which diverges from the pattern of results in the other conditions, and across our studies. Future research should further examine whether the perspective and framing of one's moral claims moderates the costs of hypocrisy (and the relative benefits of absolute proclamations).

## Table SI-1

*Full Descriptive Statistics for Study 1*

| **Panel A. Hypocrisy** | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| **Perspective** | | **Frame** | | | | | | | | |
| | | Lie | | | Truth | | | Across Frames | | |
| | | **Proclamation** | | | | | | | | |
| | | Absolute | Flexible | *Total* | Absolute | Flexible | *Total* | Absolute | Flexible | *Total* |
| **Third Person** | *M* | 5.48 | 4.74 | *5.16* | 5.36 | 4.76 | *5.06* | 5.42 | 4.76 | *5.11* |
| | *SD* | 1.22 | 1.29 | *1.30* | 1.24 | 1.35 | *1.33* | 1.23 | 1.32 | *1.31* |
| | *n* | 83 | 63 | *146* | 88 | 90 | *178* | 171 | 153 | *324* |
| **First Person** | *M* | 5.42 | 4.35 | *4.98* | 5.27 | 4.74 | *4.99* | 5.35 | 4.57 | *4.99* |
| | *SD* | 0.96 | 1.25 | *1.21* | 1.26 | 1.24 | *1.27* | 1.10 | 1.25 | *1.24* |
| | *n* | 103 | 71 | *174* | 83 | 95 | *178* | 186 | 166 | *352* |
| **Second Person** | *M* | 5.31 | 4.95 | *5.12* | 5.44 | 4.84 | *5.15* | 5.38 | 4.90 | *5.14* |
| | *SD* | 1.09 | 1.26 | *1.20* | 1.12 | 1.19 | *1.19* | 1.11 | 1.23 | *1.19* |
| | *n* | 82 | 93 | *175* | 89 | 81 | *170* | 171 | 174 | *345* |
| ***Total*** | *M* | 5.41 | 4.71 | *5.08* | 5.36 | 4.78 | *5.07* | 5.38 | 4.75 | *5.08* |
| | *SD* | 1.08 | 1.29 | *1.23* | 1.21 | 1.26 | *1.27* | 1.14 | 1.27 | *1.25* |
| | *n* | 268 | 227 | *495* | 260 | 266 | *526* | 528 | 493 | *1021* |
| **Panel B. Morality** | | | | | | | | | | |
| **Perspective** | | **Frame** | | | | | | | | |
| | | Lie | | | Truth | | | Across Frames | | |
| | | **Proclamation** | | | | | | | | |
| | | Absolute | Flexible | *Total* | Absolute | Flexible | *Total* | Absolute | Flexible | *Total* |
| **Third Person** | *M* | 2.76 | 2.23 | *2.53* | 2.80 | 2.29 | *2.54* | 2.78 | 2.27 | *2.54* |
| | *SD* | 1.29 | 0.95 | *1.18* | 1.17 | 1.04 | *1.13* | 1.23 | 1.00 | *1.15* |
| | *n* | 83 | 63 | *146* | 88 | 90 | *178* | 171 | 153 | *324* |
| **First Person** | *M* | 2.85 | 2.65 | *2.77* | 2.83 | 2.17 | *2.48* | 2.84 | 2.38 | *2.62* |
| | *SD* | 1.27 | 1.19 | *1.24* | 1.20 | 0.83 | *1.07* | 1.24 | 1.03 | *1.16* |
| | *n* | 103 | 71 | *174* | 83 | 95 | *178* | 186 | 166 | *352* |
| **Second Person** | *M* | 3.07 | 2.30 | *2.66* | 2.90 | 2.30 | *2.62* | 2.98 | 2.30 | *2.64* |
| | *SD* | 1.27 | 1.06 | *1.22* | 1.18 | 1.04 | *1.15* | 1.22 | 1.05 | *1.19* |
| | *n* | 82 | 93 | *175* | 89 | 81 | *170* | 171 | 174 | *345* |
| ***Total*** | *M* | 2.89 | 2.39 | *2.66* | 2.84 | 2.25 | *2.54* | 2.87 | 2.32 | *2.60* |
| | *SD* | 1.28 | 1.18 | *1.22* | 1.09 | 0.97 | *1.12* | 1.23 | 1.03 | *1.17* |
| | *n* | 268 | 260 | *495* | 227 | 266 | *526* | 528 | 493 | *1021* |

**Panel C. Future Honesty**

| **Perspective** | | **Frame** | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | Lie | | | Truth | | | Across Frames | | |
| | | **Proclamation** | | | | | | | | |
| | | Absolute | Flexible | *Total* | Absolute | Flexible | *Total* | Absolute | Flexible | *Total* |
| **Third Person** | *M* | 3.25 | 2.43 | *2.90* | 3.01 | 2.52 | *2.76* | 3.13 | 2.48 | *2.82* |
| | *SD* | 1.15 | 0.78 | *1.08* | 1.05 | 0.92 | *1.01* | 1.10 | 0.86 | *1.05* |
| | *n* | 83 | 63 | *146* | 88 | 90 | *178* | 171 | 153 | *324* |
| **First Person** | *M* | 3.13 | 2.68 | *2.95* | 2.98 | 2.31 | *2.62* | 3.06 | 2.47 | *2.78* |
| | *SD* | 1.00 | 0.81 | *0.95* | 1.11 | 0.82 | *1.02* | 1.05 | 0.84 | *1.00* |
| | *n* | 103 | 71 | *174* | 83 | 95 | *178* | 186 | 166 | *352* |
| **Second Person** | *M* | 3.23 | 2.41 | *2.80* | 3.11 | 2.43 | *2.78* | 3.17 | 2.42 | *2.79* |
| | *SD* | 1.07 | 0.90 | *1.06* | 1.02 | 0.96 | *1.05* | 1.04 | 0.92 | *1.05* |
| | *n* | 82 | 93 | *175* | 89 | 81 | *170* | 171 | 174 | *345* |
| ***Total*** | *M* | 3.20 | 2.50 | *2.88* | 3.03 | 2.42 | *2.72* | 3.12 | 2.45 | *2.80* |
| | *SD* | 1.06 | 0.85 | *1.03* | 1.06 | 0.90 | *1.03* | 1.06 | 0.87 | *1.03* |
| | *n* | 268 | 227 | *495* | 260 | 266 | *526* | 528 | 493 | *1021* |

**Panel D. Voting Intentions**

| **Perspective** | | **Frame** | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | Lie | | | Truth | | | Across Frames | | |
| | | **Proclamation** | | | | | | | | |
| | | Absolute | Flexible | *Total* | Absolute | Flexible | *Total* | Absolute | Flexible | *Total* |
| **Third Person** | *M* | 3.04 | 2.17 | *2.66* | 3.11 | 2.46 | *2.78* | 3.08 | 2.34 | *2.73* |
| | *SD* | 1.48 | 1.20 | *1.43* | 1.46 | 1.25 | *1.39* | 1.46 | 3.08 | *1.41* |
| | *n* | 83 | 63 | *146* | 88 | 90 | *178* | 171 | 153 | *324* |
| **First Person** | *M* | 2.91 | 2.69 | *2.82* | 3.06 | 2.31 | *2.66* | 2.98 | 2.47 | *2.74* |
| | *SD* | 1.28 | 1.48 | *1.36* | 1.49 | 1.16 | *1.37* | 1.38 | 2.98 | *1.37* |
| | *n* | 103 | 71 | *174* | 83 | 95 | *178* | 186 | 166 | *352* |
| **Second Person** | *M* | 3.00 | 2.16 | *2.55* | 2.98 | 2.40 | *2.70* | 2.99 | 2.27 | *2.63* |
| | *SD* | 1.38 | 1.12 | *1.31* | 1.50 | 1.15 | *1.37* | 1.44 | 2.99 | *1.34* |
| | *n* | 82 | 93 | *175* | 89 | 81 | *170* | 171 | 174 | *345* |
| *Total* | *M* | 2.98 | 2.33 | *2.68* | 3.05 | 2.38 | *2.71* | 3.01 | 2.36 | *2.70* |
| | *SD* | 1.37 | 1.48 | *1.37* | 1.28 | 1.18 | *1.38* | 1.42 | 1.23 | *1.37* |
| | *n* | 268 | 227 | *495* | 260 | 266 | *526* | 528 | 493 | *1021* |

**Multiple Mediation Additional Results**

In addition to the mediation analysis on moral evaluations reported in the main manuscript, we conducted a second mediation analysis on voting intentions. Using a bootstrapped procedure with the lavaan package in R (Rosseel, 2012), proclamation was entered as the independent variable (0 = absolute, 1 = flexible), which we collapsed across perspective and frame given the null effects in the ANOVAs. Hypocrisy and future honesty were entered as simultaneous mediators and voting intentions was entered as the dependent variable. Results indicate a significant indirect influence through hypocrisy (Indirect Effect through Hypocrisy = 0.161, 95% CI = [0.11, 0.22]) and through future honesty (Indirect Effect through Future Honesty = -0.433, 95% CI = [-0.53, -0.35]), similar to the results in the model on moral evaluations. The indirect effect through future honesty was much larger than the indirect effect through hypocrisy.

## 1.2 Study 2 Additional Results

In Study 2, in addition to the multiple mediation analysis examining the relationship between proclamation and moral evaluations reported in the main manuscript, we also conducted a second multiple mediation analysis to examine the relationship between proclamation and voting intentions.

Proclamation was entered as the independent variable (0 = absolute, 1 = flexible) and T2 voting intentions as the dependent variable. Hypocrisy (which was only measured at T2) and T2 evaluations of future honesty were entered as simultaneous mediators. As in the main manuscript, we excluded participants in the control condition in this analysis, since we did not have theoretical predictions regarding mechanisms in the control condition.

Results indicate a significant indirect influence through hypocrisy (Indirect Effect through Hypocrisy = 0.181, 95% CI = [0.09, 0.29]) and future honesty (Indirect Effect through Future Honesty = -0.393, 95% CI = [-0.57, -0.25]). As in the moral evaluations model, the flexible proclamation lowered perceptions of both hypocrisy and future honest behavior, which had opposite effects on voting intentions.

### 1.3 Study 3 Additional Results

We ran a series of mixed ANOVAs to investigate the effects of proclamation (absolute vs. flexible), order (proclamation first vs. behavior first), and time (T1 vs. T2) on each attitudinal variable of interest. We focused on the detailed results for the morality model in the main manuscript but provide extended results for the hypocrisy and future honesty models in this section. Descriptive statistics for the attitudinal measures are included in Table SI-2. We also report results from additional measures not included in the main manuscript.

**Hypocrisy**

There was a main effect of proclamation on hypocrisy ($F(1, 596) = 5.35$, $p = .021$, $\eta_p^2 < .01$), such that communicators who endorsed the absolute proclamation were viewed as more hypocritical than communicators who endorsed the flexible proclamation. There was also a main effect of time ($F(1, 596) = 43.70$, $p < .001$, $\eta_p^2 = .07$); communicators were seen as more hypocritical at T2 compared to T1.

These effects were qualified by a significant proclamation x time interaction ($F(1, 596) = 95.40$, $p < .001$, $\eta_p^2 = .14$). As reported in the main manuscript, communicators who endorsed the absolute proclamation were seen as more hypocritical at T2 than T1 ($t(305) = -10.09$, $p < .001$, $d_{RM} = 0.58$), whereas communicators who endorsed the flexible proclamation were seen as *less* hypocritical at T2 than T1 ($t(293) = 2.78$, $p = .006$, $d_{RM} = 0.16$). Participants noticed the

consistency or inconsistency in the communicator's proclamation and behavior and adjusted their hypocrisy ratings accordingly.

There was also a significant proclamation x order interaction, ($F(1, 596) = 7.76$, $p = .006$, $\eta_p^2 = .01$). The order in which information was presented significantly influenced evaluations of communicators who endorsed the absolute proclamation. Those communicators appeared more hypocritical when they endorsed an honesty proclamation after (vs. before) playing the deception game ($t(610) = 2.85$, $p = .005$, $d = 0.23$), consistent with prior research (Barden et al., 2005). However, order did not significantly influence perceptions of communicators who endorsed the flexible proclamation ($t(586) = -1.67$, $p = .095$, $d = -0.14$) — they were evaluated similarly regardless of whether the proclamation or behavior was revealed first.

No other main effects or interactions were significant at the $p < .05$ level ($ps > .104$).

**Future Honesty**

There was a main effect of proclamation on future honesty ($F(1, 596) = 220.75$, $p < .001$, $\eta_p^2 = .27$). Communicators who endorsed the absolute proclamation were seen as more committed to future honesty and more likely to engage in future honesty than communicators who endorsed the flexible proclamation. There was also a significant main effect of order ($F(1, 596) = 44.37$, $p < .001$, $\eta_p^2 = .07$), such that communicators were seen as more likely to engage in future honesty when the proclamation (vs. behavior) was presented first. Additionally, there was a main effect of time ($F(1, 596) = 50.74$, $p < .001$, $\eta_p^2 = .08$). Communicators were seen as more likely to engage in future honesty when evaluated at T1 versus T2.

These effects were qualified by a significant proclamation x time interaction ($F(1, 596) = 13.71$, $p < .001$, $\eta_p^2 = .02$). All communicators were seen as less likely to engage in future honesty when evaluated at T2 versus T1. In contrast to moral evaluations, even communicators

who endorsed the flexible proclamation were rated as less committed to future honesty when evaluated at T2. As mentioned in the main manuscript, the penalty in future honesty ratings over time was smaller for communicators who endorsed the flexible proclamation compared to communicators who endorsed the absolute proclamation.

There was a significant proclamation x order interaction ($F(1, 596) = 71.14$, $p < .001$, $\eta_p^2 = .11$). Communicators who endorsed the absolute proclamation were viewed as more likely to engage in future honesty when the proclamation was presented before (vs. after) their behavior ($t(610) = -10.46$, $p < .001$, $d = -0.85$), but order did not significantly influence inferences of future honesty among communicators who endorsed the flexible proclamation ($t(586) = -1.67$, $p = .096$, $d = 0.14$).

There was also a significant order x time interaction ($F(1, 596) = 100.96$, $p < .001$ $\eta_p^2 = .14$). When the proclamation was revealed before the behavior, communicators were seen as more likely to engage in future honesty when evaluated at T1 than T2 ($t(304) = 9.65$, $p < .001$, $d_{RM} = 0.55$); however, when the behavior was presented before the proclamation, communicators were seen as more likely to engage in future honesty when evaluated at T2 than T1 ($t(294) = -2.24$, $p = .026$, $d_{RM} = -0.13$). Thus, expectations of future honesty were always higher after the proclamation than the behavior, regardless of time point.

These effects were further qualified by a significant proclamation x order x time interaction ($F(1, 596) = 164.18$, $p < .001$, $\eta_p^2 = .22$; see Figure 3 in main manuscript). Communicators who endorsed the absolute proclamation were viewed as very committed to honesty and likely to engage in future honesty initially (T1 after the proclamation), but hypocrisy detracted from these inferences over time. Communicators who endorsed the flexible proclamation, on the other hand, did not experience as strong a reduction in future honesty

ratings over time. Nevertheless, communicators endorsing the absolute proclamation were still seen as more likely to engage in future honesty than communicators who endorsed the flexible proclamation at T2.

**Table SI-2**

*Descriptive Statistics in Study 3*

| **Panel A. Hypocrisy** | | | | |
|---|---|---|---|---|
| **Time** | | **Proclamation** | | |
| **Time One (T1)** | | Absolute | Flexible | *Total* |
| | M | 3.14 | 3.13 | 3.13 |
| Behavior First | SD | 1.31 | 1.21 | 1.26 |
| | n | 148 | 147 | 295 |
| | M | 2.61 | 3.29 | 2.94 |
| Proclamation First | SD | 1.11 | 1.07 | 1.15 |
| | n | 158 | 147 | 305 |
| | M | 2.86 | 3.21 | 3.03 |
| *Total* | SD | 1.24 | 1.15 | 1.21 |
| | n | 306 | 294 | 600 |
| **Time Two (T2)** | | | | |
| | M | 3.88 | 2.95 | 3.42 |
| Behavior First | SD | 1.59 | 1.27 | 1.51 |
| | n | 148 | 147 | 295 |
| | M | 3.71 | 3.11 | 3.42 |
| Proclamation First | SD | 1.65 | 1.19 | 1.47 |
| | n | 158 | 147 | 305 |
| | M | 3.79 | 3.03 | 3.42 |
| *Total* | SD | 1.62 | 1.23 | 1.49 |
| | n | 306 | 294 | 600 |
| **Panel B. Morality** | | | | |
| **Time** | | **Proclamation** | | |
| **Time One (T1)** | | Absolute | Flexible | *Total* |
| | M | 4.22 | 4.20 | 4.21 |
| Behavior First | SD | 1.51 | 1.41 | 1.46 |
| | n | 148 | 147 | 295 |
| | M | 5.78 | 3.64 | 4.75 |
| Proclamation First | SD | 0.84 | 1.13 | 1.46 |
| | n | 158 | 147 | 305 |
| | M | 5.02 | 3.92 | 4.48 |
| *Total* | SD | 1.44 | 1.31 | 1.48 |
| | n | 306 | 294 | 600 |

| **Panel B. Morality (cont'd)** | | | | |
|---|---|---|---|---|
| **Time** | | **Proclamation** | | |
| **Time two (T2)** | | Absolute | Flexible | *Total* |
| | M | 4.33 | 3.90 | 4.12 |
| Behavior First | SD | 1.51 | 1.40 | 1.47 |
| | n | 148 | 147 | 295 |
| | M | 4.74 | 3.94 | 4.35 |
| Proclamation First | SD | 1.65 | 1.39 | 1.58 |
| | n | 158 | 147 | 305 |
| | M | 4.54 | 3.92 | 4.24 |
| *Total* | SD | 1.59 | 1.39 | 1.53 |
| | n | 306 | 294 | 600 |
| **Panel C. Future Honesty** | | | | |
| **Time** | | **Proclamation** | | |
| **Time one (T1)** | | Absolute | Flexible | *Total* |
| | M | 3.85 | 3.80 | 3.83 |
| Behavior First | SD | 1.03 | 1.07 | 1.05 |
| | n | 148 | 147 | 295 |
| | M | 6.03 | 3.54 | 4.83 |
| Proclamation First | SD | 0.83 | 0.81 | 1.49 |
| | n | 158 | 147 | 305 |
| | M | 4.98 | 3.67 | 4.34 |
| *Total* | SD | 1.44 | 0.95 | 1.39 |
| | n | 306 | 294 | 600 |
| **Time two (T2)** | | | | |
| | M | 4.41 | 3.51 | 3.97 |
| Behavior First | SD | 1.39 | 0.90 | 1.25 |
| | n | 148 | 147 | 295 |
| | M | 4.46 | 3.51 | 4.00 |
| Proclamation First | SD | 1.37 | 1.00 | 1.29 |
| | n | 158 | 147 | 305 |
| | M | 4.44 | 3.51 | 3.98 |
| *Total* | SD | 1.38 | 0.95 | 1.27 |
| | n | 306 | 294 | 600 |

## Additional Measures and Results

In Study 3, participants played a game called The Choice Game and answered a series of questions about perceived moral standards, realism, and guilt of the communicator, which were not reported in the main manuscript. We include information about these measures in Tables SI-3 and SI-4.

### *The Choice Game*

Participants were assigned to be Player 1 in The Choice Game and the communicator was assigned to be Player 2. The game is a variant of the trust game (Berg et al., 1995). Participants read the following instructions to learn about the game (they also had to answer comprehension check questions about these instructions before progressing to the game; participants who failed two comprehension check questions were automatically excluded from progressing):

**Figure SI-1**

*Instructions for The Choice Game*

**The Choice Game Instructions**

**Please read these instructions carefully. You will be asked comprehension questions about the rules of this exercise.**

**The Choice Game Instructions: Player 1**

Player 1 starts with $2.

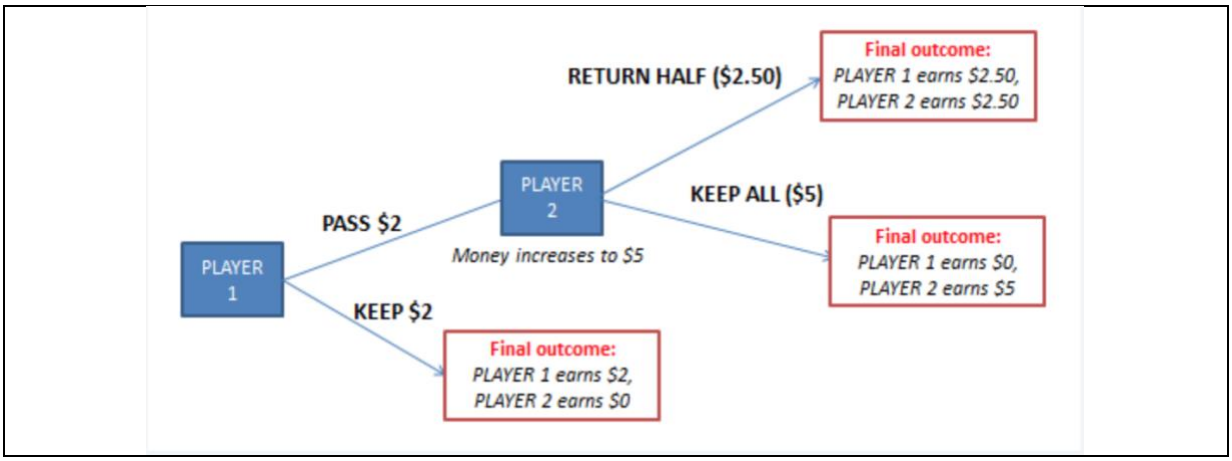
Player 1 can choose "KEEP $2" or "PASS $2."

If Player 1 chooses "KEEP $2," Player 1 earns a $2 bonus for this exercise and Player 2 earns $0 for this exercise.

If Player 1 chooses "PASS $2," the amount of money increases to $5, and Player 2 makes a decision.

If Player 1 chooses "PASS $2," Player 2 can choose "RETURN HALF ($2.50)" or "KEEP ALL ($5)." If Player 2 chooses "RETURN HALF ($2.50)," both Player 1 and Player 2 earn $2.50. If Player 2 chooses "KEEP ALL ($5)," Player 1 earns $0 and Player 2 earns $5.

See the decision tree below for an illustration.




A logistic regression was conducted to determine the effects of proclamation (absolute vs. flexible), order (proclamation first vs. behavior first), and their interaction on the decision to keep $2 or pass $2 (keep = 0, pass = 1) to Player 2. In this pre-registered model, neither the main effect of proclamation, order, or the proclamation x order interaction were significant at the $p <$ .05 level (*ps* > .146).

We also ran an exploratory model that was not pre-registered to examine the main effects of proclamation and order on trust, with no interaction term. In this model, there was a main effect of proclamation ($b$ = 0.33, $p$ = .048), providing some evidence that proclamations may impact trust, such that participants may be slightly more likely to pass to communicators making the absolute over the flexible proclamation, consistent with attitudinal evaluations of morality and future honesty.

**Table SI-3**

*Results from The Choice Game in Study 3*

| **Variable** | **Trust** | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | Model 1 | | | | Model 2 | | | |
| | *b* | SE | *z* | *p* | *b* | SE | *z* | *p* |
| Proclamation | 0.33 | 0.16 | 1.98 | .048 | 0.34 | 0.23 | 1.46 | .146 |
| Order | -0.07 | 0.16 | -0.43 | .668 | -0.06 | 0.23 | -0.24 | .814 |
| Proclamation x Order | | | | | -0.03 | 0.33 | -0.09 | .928 |

*Note*. Coding is as follows: proclamation: absolute = 1, flexible = 0; order: proclamation first = 1, behavior first = 0; outcome: pass = 1, keep = 0.

***Moral Standards***

In addition to assessing moral evaluations, we also measured perceived moral standards of the communicator in order to further investigate the moral signal of absolute proclamations. Participants answered the following question, "This person has high moral standards," (1 = Strongly disagree to 7 = Strongly agree). We did not include this item in our composite variable of morality in order to maintain consistency across our studies. We conducted a mixed ANOVA to examine the effects of proclamation (absolute vs. flexible), order (proclamation first vs. behavior first) and time (T1 vs. T2) on moral standards.

There was a main effect of proclamation ($F(1, 596) = 89.88$, $p < .001$, $\eta_p^2 = .13$), such that communicators who endorsed the absolute proclamation were seen as having higher moral standards than communicators who endorsed the flexible proclamation. There was a main effect of order ($F(1, 596) = 7.68$, $p = .006$, $\eta_p^2 = .01$), such that communicators were rated as having higher moral standards when the proclamation was presented before the behavior. There was also a main effect of time ($F(1, 596) = 16.57$, $p < .001$, $\eta_p^2 = .03$), such that communicators were rated as having lower moral standards at T2 than T1.

These effects were further qualified by a significant proclamation x order interaction ($F(1, 596) = 50.23$, $p < .001$, $\eta_p^2 = .08$). Communicators who endorsed the absolute proclamation were seen as having higher moral standards when the proclamation (vs. behavior) was revealed first ($t(610) = -8.24$, $p < .001$, $d = -0.67$). The reverse pattern occurred for communicators who endorsed the flexible proclamation. These communicators were seen as having higher moral standards when the behavior (vs. proclamation) was revealed first ($t(586) = 3.75$, $p < .001$, $d = 0.31$).

There was also a significant proclamation x time interaction ($F(1, 596) = 23.87$, $p < .001$, $\eta_p^2 = .04$). Communicators who endorsed the absolute proclamation were seen as having higher

moral standards at T1 than T2 ($t(305) = 5.91$, $p < .001$, $d_{RM} = 0.34$). Communicators who endorsed the flexible proclamation were evaluated similarly at both time points ($t(293) = -0.58$, $p = .562$, $d_{RM} = -0.03$).

Additionally, there was a significant three-way proclamation x order x time interaction ($F(1, 596) = 68.39$, $p < .001$, $\eta_p^2 = .10$). In line with moral evaluations, the absolute proclamation reflected positively on moral standards, despite the reveal of deception, but the discrepancy in moral standard evaluations of communicators endorsing absolute versus flexible honesty was largest at T1 when the proclamation was revealed first. Thus, participants did adjust their perceptions of moral standards of communicators who endorsed the absolute proclamation and lied (resulting in lower moral standards scores over time), consistent with our other results. However, these communicators were still seen as having higher moral standards overall than those who endorsed the flexible proclamation.

The order x time interaction was not significant at the $p < .05$ level ($ps > .093$).

***Realism***

Although endorsing the absolute proclamation generally reflected positively on communicators in these studies, we were curious to see if endorsing the flexible proclamation (i.e., "*It is sometimes okay for people to lie*") might be rewarded for seeming more realistic. Therefore, we asked participants to answer the question, "This person is realistic," (1 = Strongly disagree to 7 = Strongly agree). We conducted a mixed ANOVA to examine the effects of proclamation (absolute vs. flexible), order (proclamation first vs. behavior first), and time (T1 vs. T2) on realism.

Results revealed only a significant three-way proclamation x order x time interaction ($F(1, 596) = 4.16$, $p = .042$, $\eta_p^2 < .01$). No other main effects or interactions were significant at

the $p < .05$ level ($ps > .523$). The three-way interaction appears to be driven by slightly higher judgments of realism for communicators endorsing the flexible proclamation over the absolute proclamation at T2 when the lie is revealed first, but a reverse pattern for communicators endorsing the absolute proclamation at T2 when the proclamation is revealed first (see Table SI-4).

***Guilt***

To assess guilt, we asked participants to answer these two questions: "This person felt uncomfortable about sending a dishonest message in The Coin Flip Game," and "This person felt guilty about sending a dishonest message in The Coin Flip Game," (1 = Strongly disagree to 7 = Strongly agree). These items were highly correlated at both time points ($r(598) = 0.93$ at T1; $r(598) = 0.91$ at T2); however, the T2 ratings were the only relevant rating for analysis because some of the participants at T1 had not yet learned about The Coin Flip Game. Therefore, we conducted an ANOVA examining the influence of proclamation, order, and the proclamation x order interaction on a composite of perceived guilt at T2 only.

There was a main effect of proclamation ($F(1, 596) = 117.01$, $p < .001$, $\eta_p^2 = .16$), such that participants inferred greater guilt from communicators who endorsed the absolute proclamation than those who endorsed the flexible proclamation. The main effect of order and the proclamation x order interaction were not significant at the $p < .05$ level ($ps > .150$). Table SI-4 presents descriptive statistics of moral standards, realism, and guilt.

**Table SI-4**

*Descriptive Statistics for Additional Attitudinal Measures in Study 3*

| **Panel A. Moral Standards** | | | | |
|---|---|---|---|---|
| **Time** | | **Proclamation** | | |
| **Time One (T1)** | | Absolute | Flexible | *Total* |
| | M | 4.42 | 4.37 | 4.40 |
| Behavior First | SD | 1.41 | 1.42 | 1.42 |
| | n | 148 | 147 | 295 |
| | M | 5.90 | 3.61 | 4.79 |
| Proclamation First | SD | 0.90 | 1.18 | 1.55 |
| | n | 158 | 147 | 305 |
| | M | 5.18 | 3.99 | 4.60 |
| *Total* | SD | 1.39 | 1.36 | 1.50 |
| | n | 306 | 294 | 600 |
| **Time Two (T2)** | | | | |
| | M | 4.47 | 4.06 | 4.27 |
| Behavior First | SD | 1.54 | 1.29 | 1.43 |
| | n | 148 | 147 | 295 |
| | M | 4.87 | 4.01 | 4.46 |
| Proclamation First | SD | 1.52 | 1.37 | 1.51 |
| | n | 158 | 147 | 305 |
| | M | 4.68 | 4.03 | 4.36 |
| *Total* | SD | 1.54 | 1.33 | 1.48 |
| | n | 306 | 294 | 600 |
| **Panel B. Realism** | | | | |
| **Time** | | **Proclamation** | | |
| **Time One (T1)** | | Absolute | Flexible | *Total* |
| | M | 5.03 | 5.01 | 5.02 |
| Behavior First | SD | 1.21 | 1.33 | 1.27 |
| | n | 148 | 147 | 295 |
| | M | 4.97 | 5.07 | 5.02 |
| Proclamation First | SD | 1.42 | 1.26 | 1.34 |
| | n | 158 | 147 | 305 |
| | M | 5.00 | 5.04 | 5.02 |
| *Total* | SD | 1.32 | 1.30 | 1.31 |
| | n | 306 | 294 | 600 |

| **Panel B. Realism (cont'd)** | | | | |
|---|---|---|---|---|
| **Time** | | **Proclamation** | | |
| **Time Two (T2)** | | Absolute | Flexible | *Total* |
| Behavior First | M | 4.91 | 5.14 | 5.02 |
| | SD | 1.37 | 1.32 | 1.35 |
| | n | 148 | 147 | 295 |
| Proclamation First | M | 5.10 | 5.02 | 5.06 |
| | SD | 1.43 | 1.36 | 1.40 |
| | n | 158 | 147 | 305 |
| *Total* | M | 5.01 | 5.08 | 5.04 |
| | SD | 1.40 | 1.34 | 1.37 |
| | n | 306 | 294 | 600 |
| **Panel C. Guilt** | | | | |
| **Time** | | **Proclamation** | | |
| **Time Two (T2)** | | Absolute | Flexible | *Total* |
| Behavior First | M | 3.76 | 2.41 | 3.09 |
| | SD | 1.70 | 1.23 | 1.63 |
| | n | 148 | 147 | 295 |
| Proclamation First | M | 3.95 | 2.59 | 3.29 |
| | SD | 1.74 | 1.40 | 1.73 |
| | n | 158 | 147 | 305 |
| *Total* | M | 3.86 | 2.50 | 3.19 |
| | SD | 1.72 | 1.32 | 1.68 |
| | n | 306 | 294 | 600 |

### 1.4 Study 4 Additional Results

In Studies 4-5, we included a measure called attitudes towards honesty, in which participants answered the same honesty preferences question as the communicators in regard to their *own* attitudes on honesty (i.e., choosing the statement that they most agreed with from a list of choices: "*Lying is always/often/sometimes/rarely/never okay*"). Consistent with our pilot results, we find that most participants did not endorse the absolute honesty proclamation themselves, despite rewarding those who do. Therefore, we cannot explain preferences for absolute and ambitious proclamations over flexible proclamations as solely the result of a

similarity effect. However, we do find that matching preferences moderate some of our effects. Detailed results are included below.

**Attitudinal Variables and Attitudes Towards Honesty**

To test the similarity effect hypothesis, we pre-registered a 3 (proclamation: absolute vs. ambitious vs. flexible) x 2 (behavior: prosocial lie vs. selfish truth) ANOVA with attitudes towards honesty added as a covariate, with each character evaluation entered as the dependent variable (morality, hypocrisy, future honesty). Adding a standardized attitudes toward honesty variable (re-scaled by both the mean and the standard deviation) to the model did not change the pattern of results in any of our ANOVA models (see Table SI-5, results are consistent with results reported in the main manuscript without attitudes towards honesty added to the model).

**Table SI-5**

*ANOVA Results from Pre-registered Analyses with Attitudes Towards Honesty*

| **Variable** | Morality | | Hypocrisy | | Future Honesty | |
|---|---|---|---|---|---|---|
| | F | *p* | F | *p* | F | *p* |
| Proclamation | 22.15 | < .001*** | 0.11 | .895 | 76.56 | < .001*** |
| Behavior | 0.00 | .993 | 7.36 | .007** | 34.16 | < .001*** |
| Attitudes Towards Honesty | 4.84 | .028* | 0.49 | .486 | 1.88 | .171 |
| Proclamation x Behavior | 1.29 | .277 | 9.28 | < .001*** | 3.68 | .026* |

*Note*. *$p$ <.05. ** $p$ < .01. *** $p$ < .001.

To further interrogate whether participants' own attitudes towards honesty interacted with the communicators' proclamation to impact these evaluations, we also ran a series of regressions examining the relationship between a dummy-coded flexible proclamation variable (1 = flexible, 0 = absolute/ambitious) and the standardized attitudes toward honesty variable (Model 1), and their interactions (Model 2), on our attitudinal variables of interest. These regressions were not pre-registered, but our empirical results suggested that participants viewed the ambitious and absolute honesty proclamations similarly in Study 4 and, therefore, we were interested to see

whether perceptions of the flexible compared to the absolute/ambitious proclamation interacted with participants' own attitudes towards honesty to predict character evaluations. We find that the attitudes towards honesty variable did interact with the flexible proclamation variable in predicting morality and hypocrisy evaluations. Judgments of morality did not seem to depend on participants' own attitudes toward honesty when communicators also endorsed the absolute proclamation. However, as participant's own attitudes toward honesty became less strict, they viewed communicators who endorsed the flexible proclamation as more moral. Furthermore, participants with more absolute attitudes towards honesty were more lenient in their hypocrisy evaluations of communicators with shared views relative to participants who endorsed flexible honesty. Therefore, we find some evidence of preference matching. However, participants judged communicators who endorsed the absolute stance the most positively, regardless of their private views, suggesting our results are not solely driven by preference matching. We include regression results in the Table SI-6 and descriptive statistics in Table SI-7.

**Table SI-6**

*Interaction Effects on Attitudinal Variables*

| **Panel A. Morality** | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| **Independent Variable** | Model 1 | | | | Model 2 | | | |
| | *b* | SE | *t* | *p* | *b* | SE | *t* | *p* |
| Flexible proclamation | -0.74 | 0.13 | -5.79 | < .001*** | -0.77 | 0.13 | -5.97 | < .001*** |
| Attitudes towards honesty | -0.13 | 0.06 | -2.15 | .032* | -0.02 | 0.07 | -0.26 | .797 |
| Flexible proclamation x Attitudes towards honesty | | | | | -0.33 | 0.13 | -2.56 | .011* |
| **Panel B. Hypocrisy** | | | | | | | | |
| **Independent Variable** | Model 1 | | | | Model 2 | | | |
| | *b* | SE | *t* | *p* | *b* | *SE* | *t* | *p* |
| Flexible proclamation | -0.06 | 0.13 | -0.44 | .658 | -0.04 | 0.13 | -0.31 | .758 |
| Attitudes towards honesty | 0.05 | 0.06 | 0.83 | .405 | -0.04 | 0.07 | -0.51 | .613 |
| Flexible proclamation x Attitudes towards honesty | | | | | 0.25 | 0.13 | 2.03 | .043* |
| **Panel C. Future Honesty** | | | | | | | | |
| **Independent Variable** | Model 1 | | | | Model 2 | | | |
| | *b* | SE | *t* | *p* | *b* | *SE* | *t* | *p* |
| Flexible proclamation | -1.16 | 0.10 | -11.14 | < .001*** | -1.17 | 0.10 | -11.24 | < .001*** |
| Attitudes towards honesty | 0.06 | 0.05 | 1.22 | . 224 | 0.12 | 0.06 | 1.95 | .051 |
| Flexible proclamation x Attitudes towards honesty | | | | | -0.17 | 0.10 | -1.65 | .010** |

*Notes.* We report the main effects and flexible proclamation x attitudes towards honesty interaction effects for the dependent variables of morality (Panel A), hypocrisy (Panel B), and future honesty (Panel C). *$p$ <.05. ** $p$ < .01. *** $p$ < .001 for coefficients in these regressions.

**Table SI-7**

*Descriptive Statistics by Proclamation Condition and Participants' Attitudes Towards Honesty*

| Variable | Proclamation | Participants' Own Attitudes Towards Honesty | | | | |
|---|---|---|---|---|---|---|
| | | *Always* | *Often* | *Sometimes* | *Rarely* | *Never* |
| Morality | Absolute | NA (NA) | 4.83 (0.83) | 4.93 (1.31) | 5.06 (1.54) | **5.14 (1.41)** |
| | Ambitious | 4.00 (NA) | 1.00 (NA) | 4.76 (1.44) | **4.64 (1.50)** | 4.06 (1.86) |
| | Flexible | 4.00 (NA) | 4.60 (0.64) | **4.55 (1.40)** | 3.79 (1.42) | 3.64 (1.27) |
| Hypocrisy | Absolute | NA (NA) | 3.05 (1.69) | 3.19 (1.48) | 2.80 (1.59) | **3.05 (1.78)** |
| | Ambitious | 4.00 (NA) | 7.00 (NA) | 2.98 (1.40) | **2.91 (1.45)** | 3.30 (1.72) |
| | Flexible | 3.60 (NA) | 3.68 (1.92) | **2.58 (1.06)** | 3.00 (1.25) | 3.59 (1.19) |
| Future honesty | Absolute | NA (NA) | 4.31 (1.14) | 4.80 (1.30) | 5.24 (1.25) | **5.16 (1.20)** |
| | Ambitious | 4.00 (NA) | 2.25 (NA) | 4.43 (1.09) | **4.84 (1.23)** | 4.26 (1.74) |
| | Flexible | 2.00 (NA) | 3.90 (0.49) | **3.85 (1.00)** | 3.49 (0.94) | 3.81 (0.95) |
| *Total number of participants* | | 2 | 10 | 195 | 297 | 94 |

*Notes.* We present means for each condition with standard deviations in parentheses. The majority of participants chose "sometimes" or "rarely" when asked to choose the following statement they most agreed with from the options, "*Lying is always/often/sometimes/rarely/never okay,*" and thus, cell totals are small for "always" and "often," leading to some NA results. We include the total participant number by attitudes towards honesty type (*always/often/sometimes/rarely/never*) in the final row. Bolded numbers indicate cells in which participants' own attitudes towards honesty and communicator proclamations match.

## Trust and Attitudes Towards Honesty

We also regressed trust behavior (the decision to rely or verify in the RV game with 1 = rely, 0 = verify) on a flexible proclamation dummy variable (1 = flexible, 0 = absolute/ambitious) and the standardized attitudes towards honesty variable (Model 1), and their interaction (Model 2). Results are reported in Table SI-8. There was a significant main effect of the flexible proclamation on trust ($b$ = -0.72, $p$ < .001), such that participants were less likely to trust communicators who endorsed the flexible versus the absolute and ambitious proclamations. The main effect of attitudes towards honesty was not significant ($p$ = .702).

These results were qualified by a significant flexible proclamation x attitudes towards honesty interaction on trust ($b$ = -0.40, $p$ = .033), such that participants who endorsed the flexible honesty stance themselves exhibited greater trust in communicators who also endorsed the flexible stance relative to participants who endorsed the absolute or ambitious stance, providing some evidence for similarity-matching on trust in Study 4. Notably, however, even participants who endorsed flexible honesty exhibited comparable trust in communicators who made flexible, ambitious, and absolute proclamations; these participants trusted communicators who endorsed the flexible stance to a greater degree than participants who endorsed absolute or ambitious honesty, but they did not exhibit *greater* levels of trust in communicators endorsing flexible honesty compared to communicators endorsing ambitious or absolute honesty, despite sharing these communicators' views. We include descriptive statistics in Table SI-9.

**Table SI-8**

*Trust Regression Results*

| **Variable** | **Trust** | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | Model 1 | | | | Model 2 | | | |
| | *b* | SE | *z* | *p* | *b* | SE | *z* | *p* |
| Flexible proclamation | -0.72 | 0.19 | -3.88 | < .001*** | -0.76 | 0.19 | -4.02 | < .001*** |
| Attitudes towards honesty | 0.03 | 0.09 | 0.38 | .702 | 0.15 | 0.10 | 1.49 | .135 |
| Flexible proclamation x Attitudes towards honesty | | | | | -0.40 | 0.19 | -2.14 | .033* |

*Notes.* We report Model 1 (the main effects of the flexible proclamation and attitudes towards honesty interaction effects) and Model 2 (the model considering the flexible proclamation x attitudes towards honesty interaction effect) for a logistic regression on trust. *$p$ <.05. ** $p$ < .01. *** $p$ < .001 for coefficients in these regressions.

**Table SI-9**

*Trust (% who Rely) Based on Proclamation Condition and Own Attitudes Towards Honesty*

| **Proclamation** | **Participants' Own Attitudes Towards Honesty** | | | | |
|---|---|---|---|---|---|
| | *Always* | *Often* | *Sometimes* | *Rarely* | *Never* |
| Absolute | *0/0* | *3/4* | *22/59* | *44/90* | ***20/39*** |
| | | 75% | 37% | 49% | **51%** |
| Ambitious | *0/1* | *0/1* | *23/61* | ***55/108*** | *13/31* |
| | 0% | 0% | 37% | **51%** | 42% |
| Flexible | *0/1* | *1/5* | ***28/75*** | *26/99* | *4/24* |
| | 0% | 20% | **37%** | 26% | 16% |
| *Cell sample size* | 2 | 10 | 195 | 297 | 94 |

*Notes.* The majority of participants chose "sometimes" or "rarely" when asked to choose the following statement they most agreed with from the options, "*Lying is always/often/sometimes/rarely/never okay.*" We list the count and the percentage of participants who rely (v. verify) on communicators who indicated that "*Lying is never/rarely/sometimes okay*," by proclamation condition (never vs. rarely vs. sometimes) and participant attitude choice (with the remaining participants per cell choosing to verify). The last three columns of this table are of greater interest due to the small number of participants choosing "always" or "often." We include the total participant numbers endorsing each stance in the final row. Bolded numbers indicate cells in which participants' own attitudes towards honesty and communicator proclamations match.

## Additional Moderated Mediation Analysis

Using Hayes' PROCESS Macro for SPSS, Model 7 (Hayes, 2017) a bootstrapped moderated mediation analysis was conducted to examine the role of hypocrisy and future honesty in mediating the relationship between a flexible proclamation dummy (1 = flexible, 0 = absolute/ambitious) and moral evaluations. Behavior was entered as the moderator (1 = prosocial lie, 0 = selfish truth).

When the communicator told a prosocial lie, we find evidence that hypocrisy and future honesty mediate the effect of proclamation type on moral evaluations (Indirect Effect through Hypocrisy = 0.231, 95% CI = [0.10, 0.38]; Indirect Effect through Future Honesty = -0.501, 95% CI = [-0.67, -0.34]). When the communicator told a selfish truth, we also find evidence that hypocrisy and future honesty mediate the effect of proclamation type on moral evaluations

(Indirect Effect through Hypocrisy = -0.189, 95% CI = [-0.33, -0.05]; Indirect Effect through Future Honesty = -0.763, 95% CI = [-0.95, -0.57]). However, the indirect effects of hypocrisy for the prosocial lies and selfish truths are in opposite directions. Consistent with the results in the model reported in the main manuscript on trust, we find evidence of moderated mediation through hypocrisy (95% CI around index of moderated mediation = .420, 95% CI = [0.23, 0.63]), and future honesty (95% CI around index of moderated mediation = .262, 95% CI = [0.06, 0.47]); see Table SI-10.

**Table SI-10**

*Moderated Mediation Results on Morality*

| **Mediator** | **Behavior** | | | | | | |
|---|---|---|---|---|---|---|---|
| | Prosocial Lie | | | Selfish Truth | | | **Index of Moderated Mediation** |
| | **X→M** | **M→Y** | **Indirect Effect** | **X→M** | **M→Y** | **Indirect Effect** | |
| Hypocrisy | -0.558** | -0.492* | **0.231 [0.10, 0.38]** | 0.456** | -0.306* | **-0.189 [-0.33, -0.05]** | **.420 [0.23, 0.63]** |
| Future Honesty | -0.939*** | 0.507* | **-0.501 [-0.67, -0.34]** | -1.430*** | 0.678* | **-0.763 [-0.95, -0.57]** | **.262 [0.06, 0.47]** |

*Notes.* The model that was tested included all items in the first column as simultaneous mediators, the dummy-coded flexible proclamation as the independent variable (1 = flexible, 0 = absolute/ambitious), moral evaluations as the dependent variable, and behavior as the moderator (1 = prosocial lie, 0 = selfish truth). Each "Indirect Effect" signifies the lower-level and upper-level 95% confidence intervals around the indirect effect for the corresponding mediator. Bold numbers indicate confidence intervals that do not contain zero. We ran separate linear regressions to compute the X→M and M→Y pathways. *$p$ <.05. ** $p$ < .01. *** $p$ < .001 for coefficients in these regressions.

## 1.5 Study 5 Additional Results

The same set of analyses described in SI Appendix 1.4 were conducted to explore the influence of attitudes towards honesty in moderating our results in Study 5. We also include a new model in Study 5 to study the relationship between attitudes towards honesty and the perceived social benefit of the proclamation itself.

**Attitudinal Variables and Attitudes Towards Honesty**

First, we ran the pre-registered analysis, which was a 3 (proclamation: absolute vs. ambitious vs. flexible) x 2 (behavior: prosocial truth vs. selfish lie) ANOVA with a standardized attitudes towards honesty variable (re-scaled by both the mean and the standard deviation) added to the model, with each of our character evaluation variables (morality, hypocrisy, future honesty, social benefit of the proclamation) entered as the dependent variable. The attitudes towards honesty variable was not a significant predictor of character evaluations and did not change the pattern of results in any of our ANOVA models (see Table SI-11, results are consistent with results reported in the main manuscript without attitudes towards honesty added).

**Table SI-11**

*ANOVA Results from Pre-registered Analyses with Attitudes Towards Honesty*

| **Variable** | Morality | | Hypocrisy | | Future Honesty | | Social Benefit of the Proclamation | |
|---|---|---|---|---|---|---|---|---|
| | F | *p* | F | *p* | F | *p* | F | *p* |
| Proclamation | 17.84 | < .001*** | 8.05 | < .001*** | 29.11 | < .001*** | 176.57 | < .001*** |
| Behavior | 791.07 | < .001*** | 568.25 | < .001*** | 795.16 | < .001*** | 7.84 | .005** |
| Attitudes Towards Honesty | 0.05 | .815 | 0.00 | .972 | 1.79 | .182 | 1.76 | .185 |
| Proclamation x Behavior | 6.80 | .001** | 66.86 | < .001*** | 11.81 | < .001*** | 0.74 | .477 |

*Note*. *$p$ <.05. ** $p$ < .01. *** $p$ < .001.

To further explore whether participants' own attitudes towards honesty and the proclamation interacted to impact character evaluations, we ran the same set of exploratory analysis as in Study 4 (see previous Appendix 1.4 for more details on the analytical strategy, if needed). Across these attitudinal models, we find that communicators who endorsed absolute or ambitious proclamations were consistently evaluated more positively, in terms of morality and future honesty, than communicators who endorsed flexible proclamations, despite increased perceptions of hypocrisy. Additionally, the perceived social benefit of the absolute and ambitious

proclamations was greater than the perceived social benefit for the flexible proclamation. The interaction effects in these models indicate that positive evaluations of communicators endorsing absolute and ambitious proclamations relative to flexible proclamations were enhanced when participants' views more closely aligned with these communicators' views. Participants who themselves endorsed flexible honesty exhibited greater moral ratings of the communicators endorsing flexible honesty relative to participants who endorsed ambitious or absolute honesty, providing evidence for preference matching, but even these participants viewed the communicators endorsing the ambitious proclamation as the most moral and most likely to engage in honesty. We include detailed regression results in Table SI-12 and descriptive statistics in Table SI-13.

**Table SI-12**

*Interaction Effects on Attitudinal Variables*

| **Panel A. Morality** | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| **Independent Variable** | Model 1 | | | | Model 2 | | | |
| | *b* | SE | *t* | *p* | *b* | SE | *t* | *p* |
| Flexible proclamation | -0.65 | 0.16 | -3.96 | < .001*** | -0.70 | 0.16 | -4.26 | < .001*** |
| Attitudes towards honesty | 0.01 | 0.08 | 0.15 | .882 | 0.13 | 0.09 | 1.47 | .141 |
| Flexible proclamation x Attitudes towards honesty | | | | | -0.46 | 0.17 | -2.63 | .009** |
| **Panel B. Hypocrisy** | | | | | | | | |
| **Independent Variable** | Model 1 | | | | Model 2 | | | |
| | *b* | SE | *t* | *p* | *b* | SE | *t* | *p* |
| Flexible proclamation | -0.41 | 0.16 | -2.54 | .011* | -0.38 | 0.16 | -2.31 | .021* |
| Attitudes towards honesty | -0.01 | 0.08 | -0.14 | .889 | -0.10 | 0.09 | -1.09 | .276 |
| Flexible proclamation x Attitudes towards honesty | | | | | 0.33 | 0.17 | -1.90 | .058 |
| **Panel C. Future Honesty** | | | | | | | | |
| **Independent Variable** | Model 1 | | | | Model 2 | | | |
| | *b* | SE | *t* | *p* | *b* | SE | *t* | *p* |
| Flexible proclamation | -0.73 | 0.15 | -4.96 | < .001*** | -0.77 | 0.15 | -5.26 | < .001*** |
| Attitudes towards honesty | 0.06 | 0.07 | 0.89 | .372 | 0.17 | 0.08 | 2.14 | .033* |
| Flexible proclamation x Attitudes towards honesty | | | | | -0.42 | 0.16 | -2.68 | .008** |
| **Panel D. Social Benefit of the Proclamation** | | | | | | | | |
| **Independent Variable** | Model 1 | Model 2 | | | | | | |
| | *b* | SE | *t* | *p* | *b* | SE | *t* | *p* |
| Flexible proclamation | -2.28 | 0.12 | -18.64 | < .001*** | -2.36 | 0.12 | -19.85 | < .001*** |
| Attitudes towards honesty | 0.08 | 0.06 | 1.40 | 0.161 | 0.29 | 0.06 | 4.50 | < .001*** |
| Flexible proclamation x Attitudes towards honesty | | | | | -0.81 | 0.13 | -6.37 | < .001*** |

| **Panel D. Social Benefit of the Proclamation** | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| **Independent Variable** | Model 1 | | | | Model 2 | | | |
| | *b* | SE | *t* | *p* | *b* | SE | *t* | *p* |
| Flexible proclamation | -2.28 | 0.12 | -18.64 | < .001*** | -2.36 | 0.12 | -19.85 | < .001*** |
| Attitudes towards honesty | 0.08 | 0.06 | 1.40 | 0.161 | 0.29 | 0.06 | 4.50 | < .001*** |
| Flexible proclamation x Attitudes towards honesty | | | | | -0.81 | 0.13 | -6.37 | < .001*** |

*Notes.* We reportmain effects and the flexible proclamation x attitudes towards honesty interaction effects for the dependent variables of morality (Panel A), hypocrisy (Panel B), future honesty (Panel C), and social benefit of the proclamation (Panel D). *$p$ <.05. ** $p$ < .01. *** $p$ < .001.

## Table SI-13

*Descriptive Statistics by Proclamation Condition and Participants' Attitudes Towards Honesty*

| **Variable** | **Proclamation** | **Participants' Own Attitudes Towards Honesty** | | | | |
|---|---|---|---|---|---|---|
| | | *Always* | *Often* | *Sometimes* | *Rarely* | *Never* |
| Morality | Absolute | 1.56 (0.51) | 4.08 (1.87) | 3.77 (1.92) | 4.34 (2.08) | **4.54 (2.27)** |
| | Ambitious | 4.00 (NA) | 4.22 (0.84) | 4.24 (1.85) | **4.46 (1.71)** | 3.40 (2.16) |
| | Flexible | NA (NA) | 4.00 (3.30) | **3.86 (1.58)** | 3.58 (1.62) | 2.76 (1.59) |
| Hypocrisy | Absolute | 6.60 (0.69) | 3.00 (1.05) | 4.32 (2.24) | 3.45 (2.26) | **3.29 (2.40)** |
| | Ambitious | 1.00 (NA) | 2.87 (1.27) | 3.15 (1.73) | **3.22 (1.90)** | 4.19 (1.92) |
| | Flexible | NA (NA) | 5.40 (0.28) | **2.83 (1.07)** | 3.00 (1.05) | 4.02 (1.54) |
| Future honesty | Absolute | 1.58 (0.52) | 3.63 (0.97) | 3.82 (1.83) | 4.56 (1.89) | **4.60 (2.13)** |
| | Ambitious | 5.25 (NA) | 3.58 (1.01) | 4.10 (1.55) | **4.55 (1.59)** | 3.46 (1.76) |
| | Flexible | NA (NA) | 2.75 (1.77) | **3.84 (1.31)** | 3.56 (1.33) | 2.92 (1.37) |
| Social benefit of the proclamation | Absolute | 4.17 (2.25) | 4.00 (1.47) | 4.25 (1.30) | 5.21 (1.27) | **5.85 (1.33)** |
| | Ambitious | 4.00 (NA) | 5.17 (1.15) | 4.44 (1.53) | **5.30 (1.16)** | 3.98 (2.04) |
| | Flexible | NA (NA) | 4.75 (1.06) | **3.26 (1.15)** | 2.58 (1.20) | 1.93 (1.29) |
| *Total number of participants* | | 4 | 9 | 144 | 330 | 109 |

*Notes.* We present means for each condition with standard deviations in parentheses. The majority of participants chose "sometimes" or "rarely" when asked to choose the following statement they most agreed with from the options, "*Lying is always/often/sometimes/rarely/never okay*," leading to some small cell totals and NA results for "always." We include the total participants by attitudes towards honesty in the final row. Bolded numbers indicate cells in which participants' own attitudes towards honesty and communicator proclamations match.

**Trust and Attitudes Towards Honesty**

We regressed trust behavior (the decision to rely or verify in the RV game with 1 = rely, 0 = verify) on a flexible proclamation dummy variable (1 = flexible, 0 = absolute/ambitious) and the standardized attitudes towards honesty variable (Model 1), and their interaction (Model 2). We report results in Table SI-14. In Model 1, there was a significant main effect of the flexible proclamation on trust ($b$ = -0.57, $p$ = .002), such that participants were less likely to trust communicators who endorsed the flexible versus the absolute and ambitious proclamations. There was also a main effect of attitudes towards honesty ($b$ = 0.09, $p$ = .271), such that as participants' attitudes towards honesty became less flexible, they were more likely to rely on the communicator.

However, in Model 2, the main effects of the flexible proclamation ($b$ = -0.67, $p$ = .001) and attitudes towards honesty ($b$ = 0.25, $p$ = .013) were qualified by a significant flexible proclamation x attitudes towards honesty interaction on trust ($b$ = -0.66, $p$ = .001). Participants who themselves thought lying was never or rarely okay exhibited more trust in communicators who endorsed the absolute and ambitious honesty stances, whereas participants who endorsed flexible honesty were more likely to trust communicators who also endorsed flexible honesty, adding further evidence for similarity-matching on trust. It is worth noting that even though the participants who endorsed flexible honesty were more likely to trust the communicators also endorsing flexible honesty (over those endorsing absolute or ambitious honesty), even among these participants, the overall rates of trust were relatively low – participants were more likely to verify than rely when they held flexible views on honesty. We include descriptive statistics in Table SI-15.

## Table SI-14

*Trust Regression Results*

| Variable | Trust | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | Model 1 | | | | Model 2 | | | |
| | *b* | SE | *z* | *p* | *b* | *SE* | *z* | *p* |
| Flexible proclamation | -0.57 | 0.19 | -3.09 | .002** | -0.67 | 0.19 | -3.44 | .001** |
| Attitudes towards honesty | 0.09 | 0.09 | 1.10 | .271 | 0.25 | 0.10 | 2.49 | .013* |
| Flexible proclamation x Attitudes towards honesty | | | | | -0.66 | 0.21 | -3.18 | .001** |

*Notes.* We report the main effects and flexible proclamation x attitudes towards honesty interaction effect for a logistic regression on trust. *$p$ <.05. ** $p$ < .01. *** $p$ < .001.

## Table SI-15

*Trust (% who Rely) Based on Proclamation Condition and Own Attitudes Towards Honesty*

| Proclamation | Participants' Own Attitudes Towards Honesty | | | | |
|---|---|---|---|---|---|
| | *Always* | *Often* | *Sometimes* | *Rarely* | *Never* |
| Absolute | *1/3* | *1/4* | *15/46* | *45/89* | ***33/59*** |
| | 33% | 25% | 33% | 51% | **56%** |
| Ambitious | *1/1* | *1/3* | *8/34* | ***62/130*** | *11/29* |
| | 100% | 33% | 24% | **48%** | 38% |
| Flexible | *0/0* | *1/2* | ***26/64*** | *30/111* | *4/21* |
| | 0% | 50% | **41%** | 27% | 19% |
| *Cell sample size* | 4 | 9 | 144 | 330 | 109 |

*Notes.* As in Study 4, the majority of participants chose "sometimes" or "rarely" when asked to choose the following statement they most agreed with from the options, "*Lying is always/often/sometimes/rarely/never okay*." We list both the count and the percentage of participants who rely (v. verify) on communicators who indicated that "*Lying is never/rarely/sometimes okay*," by proclamation condition (never vs. rarely vs. sometimes) and participant attitude choice (with the remaining participants per cell choosing to instead verify the communicator's message). However, the last three columns of this table are of greater interest due to the small number of participants choosing the options "always" or "often" overall. We include the total participant numbers endorsing each stance overall in the final row. Bolded numbers indicate cells in which participants' own attitudes towards honesty and communicator proclamations match.

**Additional Moderated Mediation Analysis**

Using Hayes' PROCESS Macro for SPSS, Model 7 (Hayes, 2017) a bootstrapped moderated mediation analysis was conducted to examine whether hypocrisy, future honesty, and the perceived social benefit of the proclamation mediate the relationship between the flexible proclamation dummy and moral evaluations, moderated by behavior (0 = prosocial truth, 1 = selfish lie).

When the communicator told a selfish lie, we find evidence that both future honesty (Indirect Effect through Future Honesty = -0.266, 95% CI = [-0.47, -0.07]) and hypocrisy (Indirect Effect through Hypocrisy = 0.371, 95% CI = [0.25, 0.51]) mediate the relationship between the flexible proclamation dummy and moral evaluations. Likewise, when the communicator told a prosocial truth, we again find evidence that future honesty (Indirect Effect through Future Honesty = -0.890, 95% CI = [-1.09, -0.70]) and hypocrisy (Indirect Effect through Hypocrisy = -0.183, 95% CI = [-0.27, -0.11]) mediate the effect of the flexible proclamation dummy on moral evaluations. However, the indirect effects of hypocrisy are in the opposite direction. The perceived social benefit of the proclamation does not significantly mediate the relationship between proclamation and moral evaluations when telling a selfish lie (Indirect Effect through Social Benefit = -0.024, 95% CI = [-0.14, 0.09]) or the prosocial truth (Indirect Effect through Social Benefit = -0.028, 95% CI = [-0.16, 0.10]).

We find evidence of moderated mediation through hypocrisy (index of moderated mediation = .554, 95% CI = [0.38, 0.74]) and future honesty (index of moderated mediation = .624, 95% CI = [0.34, 0.91]). We find no evidence of moderated mediation through the perceived social benefit of the proclamation (index of moderated mediation = .003, 95% CI = [-0.02, 0.02]).

**Table SI-16**

*Moderated Mediation Results on Morality*

| **Mediator** | **Behavior** | | | | | | |
|---|---|---|---|---|---|---|---|
| | Selfish Lie | | | Prosocial Truth | | | **Index of Moderated Mediation** |
| | **X→M** | **M→Y** | **Indirect Effect** | **X→M** | **M→Y** | **Indirect Effect** | |
| Hypocrisy | -1.629*** | -0.216*** | **0.371 [0.25, 0.51]** | 0.805*** | -0.305*** | **-0.183 [-0.27, -0.11]** | **.554 [0.38, 0.74]** |
| Future honesty | -0.336* | 0.710*** | **-0.266 [-0.47, -0.07]** | -1.126*** | 0.564*** | **-0.890 [-1.09, -0.70]** | **.624 [0.34, 0.91]** |
| Social benefit of the proclamation | -2.157*** | -0.014 | -0.024 [-0.14, 0.09] | -2.436*** | 0.057 | -0.028 [-0.16, 0.10] | .003 [-0.02, 0.02] |

*Notes.* The model that was tested included all items in the first column as simultaneous mediators, the dummy-coded flexible proclamation as the independent variable (1 = flexible, 0 = absolute/ambitious), moral evaluations as the dependent variable, and behavior as the moderator (1 = selfish lie, 0 = prosocial truth). Each Indirect Effect is accompanied by the 95% confidence interval around the indirect effect. Bold numbers indicate confidence intervals that do not contain zero. We ran separate linear regressions to compute the X→M and M→Y pathways. *$p$ <.05. ** $p$ < .01. *** $p$ < .001 for coefficients in these regressions.

# 2. SUPPLEMENTAL STUDIES 1-7: METHODS AND RESULTS

## 2.1 Supplemental Study 1

### Abbreviated Methods

Supplemental Study 1 compared evaluations of communicators who endorsed an absolute honesty proclamation to communicators with no known honesty proclamation (neutral control). As in Studies 3-5, communicators were past study participants who had completed a survey on beliefs and preferences in a prior study and played a deception game with another participant (who was not featured in the main study). Half of the communicators in the previous study were asked to indicate their honesty beliefs in the prior study and chose the option "*It is never okay for people to lie*" (absolute proclamation) and half of the communicators were not asked about their honesty beliefs in the survey (neutral control). All other questions in the survey on beliefs and preferences were held constant between conditions. All communicators told a lie in the deception game (the previously described Coin Flip Game), but we varied whether the lie was told for prosocial or selfish motives.

We intended to recruit 400 participants from MTurk. Participants had to answer comprehension check questions correctly throughout the survey to progress to the end of the study. Of the 399 participants who completed the survey, 37 responses were eliminated due to repeated use of a single MTURK ID or failure to provide consent, resulting in a final sample of 360 participants ($N$ =360, 45.83% female, $M_{age}$ = 35.81, $SD_{age}$ = 10.95). Participants were asked to evaluate communicators based on their survey responses and past behavior; participants were randomly assigned to one of four conditions in a 2 (proclamation: absolute vs. control) x 2 (behavior: prosocial lie vs. selfish lie) between-subjects design.

After learning about the communicator's survey responses and behavior in the previous study, participants rated the communicator on perceived morality, hypocrisy, benevolence, and guilt. Participants also played a trust game with the communicator called The Choice Game (see Appendix 1.3 for instructions of The Choice Game) and we measured two related attitudinal measures of trust. Finally, participants were asked to provide basic demographic information and explain the intended purpose of The Coin Flip Game before payment and exiting.

***Dependent Variables***

**Morality.** Participants were asked to rate the communicator on the trait "moral" (1 = Not at all to 7 = Extremely).

**Hypocrisy**. Participants indicated agreement with the following statement, "This person is a hypocrite," (1 = Strongly disagree to 7 = Strongly agree).

**Guilt.** Participants rated agreement with the statement "This person felt uncomfortable about the decision to SEND a dishonest message," (1 = Strongly disagree to 7 = Strongly agree).

**Benevolence.** Benevolence was measured by asking participants to rate communicators on three related traits: "Kind," "Benevolent," and "Selfish" [reverse-coded] (1 = Not at all to 7 = Extremely). Items were combined due to high reliability ($\alpha = .79$).

**Attitudinal Trust.** In addition to playing The Choice Game with the communicator, participants also answered the following questions regarding their beliefs about their partner's behavior in The Choice Game: "I trust Player 2 to RETURN money" and "I am confident that Player 2 will RETURN money" (1 = Strongly disagree to 7 = Strongly agree). The two items were combined into a single measure of attitudinal trust due to high reliability ($r(358) = .95$, $p < .001$).

**Results**

***Trust***

We used logistic regression to determine the effects of proclamation (1 = absolute, 0 = control) and behavior (1 = selfish lie, 0 = prosocial lie) on the decision to keep $2 or pass $2 to Player 2 (1 = pass, 0 = keep) in The Choice Game. There was a significant main effect of behavior ($b$ = -1.35, $p$ < .001), such that participants were less willing to pass $2 when communicators told a selfish lie than a prosocial lie. The proclamation x behavior interaction was not significant at the $p$ < .05 level ($p$ = .119), when the modeled was conducted with the interaction term included.

**Table SI-17**

*Results from The Choice Game in Supplemental Study 1*

| **Variable** | **Trust** | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | Model 1 | | | | Model 2 | | | |
| | *b* | SE | *z* | *p* | *b* | SE | *z* | *p* |
| Proclamation | 0.22 | 0.23 | 0.94 | .346 | -0.09 | 0.30 | -0.30 | .761 |
| Behavior | -1.36 | 0.23 | -5.88 | < .001*** | -1.76 | 0.35 | -4.98 | < .001*** |
| Proclamation x Behavior | | | | | 0.73 | 0.47 | 1.56 | .119 |

*Notes.* Proclamation: control is the baseline (0) and absolute honesty = 1. Behavior: prosocial lies are the baseline (0) and selfish lies are 1. Outcome: pass= 1, keep = 0. *$p$ <.05. ** $p$ < .01. *** $p$ < .001 for coefficients in these regressions.

***Attitudinal variables (morality, hypocrisy, guilt, benevolence, attitudinal trust)***

A series of ANOVAs were conducted to examine the effects of proclamation (absolute vs. control) and behavior (prosocial lie vs. selfish lie) on each attitudinal variable. Descriptive statistics for these measures are presented in Table SI-18.

**Morality.** There was a main effect of proclamation on moral evaluations ($F$(1, 356) = 8.99, $p$ = .003, $\eta_p^2$ = .02). Communicators who endorsed the absolute proclamation were viewed as more moral than communicators with no known stance on honesty. Additionally, there was a main effect of behavior ($F$(1, 356) = 87.83, $p$ < .001, $\eta_p^2$ = .20). Communicators who told

prosocial lies were viewed as more moral than communicators who told selfish lies. The proclamation x behavior interaction was not significant at the $p < .05$ level ($p = .916$).

**Hypocrisy**. Results revealed a main effect of proclamation on hypocrisy ($F(1, 356) = 5.40$, $p = .021$, $\eta_p^2 = .01$). Communicators who endorsed the absolute proclamation were viewed as more hypocritical than communicators with no known stance on honesty. There was also a main effect of behavior ($F(1, 356) = 90.08$, $p < .001$, $\eta_p^2 = .20$), such that communicators who told selfish lies were seen as more hypocritical than communicators who told prosocial lies. The proclamation x behavior interaction was not significant at the $p < .05$ level ($p = .130$).

**Guilt**. There was a main effect of proclamation on guilt ($F(1, 356) = 7.64$, $p = .006$, $\eta_p^2 = .02$). Participants inferred greater guilt from communicators who lied after endorsing the absolute proclamation than communicators who lied and had no known stance on honesty. There was also a main effect of behavior ($F(1, 356) = 4.34$, $p = .038$, $\eta_p^2 = .01$), such that communicators who told prosocial lies were perceived as experiencing greater guilt compared to communicators who told selfish lies. The proclamation x behavior interaction was not significant at the $p < .05$ level ($p = .232$).

**Attitudinal Trust.** There was a main effect of behavior on attitudinal trust ($F(1, 356) = 42.12$, $p < .001$, $\eta_p^2 = .11$). Communicators who told prosocial lies were trusted more than communicators who told selfish lies. Neither the main effect of proclamation nor the proclamation x behavior interaction were significant at the $p < .05$ level ($ps > .128$).

**Benevolence**. There was a main effect of proclamation on benevolence ($F(1, 356) = 6.50$, $p = .011$, $\eta_p^2 = .02$). Communicators who endorsed the absolute proclamation were seen as more benevolent than communicators with no known stance on honesty. There was also a main effect of behavior ($F(1, 356) = 162.09$, $p < .001$, $\eta_p^2 =. 31$), such that communicators who told prosocial

lies were viewed as more benevolent than communicators who told selfish lies. The proclamation x behavior interaction was not significant at the $p < .05$ level ($p = .590$).

***Mediation***

A multiple mediation analysis was conducted using a bootstrapped procedure in the lavaan package in R (Rosseel, 2012) to further test the effect of proclamation on moral evaluations. Proclamation was entered as the independent variable (1 = absolute, 0 = control), ratings of hypocrisy and guilt as simultaneous mediators, and moral evaluations as the dependent variable. We found evidence of a significant indirect effect through guilt (Indirect Effect through Guilt = 0.158, 95% CI = [0.04, 0.29]), but not hypocrisy (Indirect Effect through Hypocrisy = -0.137, 95% CI = [-0.29, 0.02]). These results suggest that participants thought communicators who endorsed the absolute proclamation incurred an emotional cost for engaging in deception, which reflected positively on their character.

**Table SI-18**

*Descriptive Statistics for Attitudinal Measures in Supplemental Study 1*

| **Panel A. Attitudinal Trust** | | | | |
|---|---|---|---|---|
| **Behavior** | | **Proclamation** | | |
| | | Absolute | Control | *Total* |
| Selfish Lie | M | 3.02 | 2.43 | 2.73 |
| | SD | 1.98 | 1.64 | 1.84 |
| | n | 92 | 89 | 181 |
| Prosocial Lie | M | 4.00 | 4.01 | 4.00 |
| | SD | 1.84 | 1.99 | 1.90 |
| | n | 95 | 84 | 179 |
| *Total* | M | 3.52 | 3.19 | 3.36 |
| | SD | 1.97 | 1.98 | 1.98 |
| | n | 187 | 173 | 360 |
| **Panel B. Benevolence** | | | | |
| **Behavior** | | **Proclamation** | | |
| | | Absolute | Control | *Total* |
| Selfish Lie | M | 3.21 | 2.87 | 3.05 |
| | SD | 1.28 | 1.14 | 1.22 |
| | n | 92 | 89 | 181 |
| Prosocial Lie | M | 4.70 | 4.49 | 4.60 |
| | SD | 1.04 | 1.16 | 1.10 |
| | n | 95 | 84 | 179 |
| *Total* | M | 3.97 | 3.66 | 3.82 |
| | SD | 1.38 | 1.40 | 1.40 |
| | n | 187 | 173 | 360 |
| **Panel C. Hypocrisy** | | | | |
| **Behavior** | | **Proclamation** | | |
| | | Absolute | Control | *Total* |
| Selfish Lie | M | 4.84 | 4.25 | 4.55 |
| | SD | 1.40 | 1.20 | 1.34 |
| | n | 92 | 89 | 181 |
| Prosocial Lie | M | 3.15 | 3.02 | 3.09 |
| | SD | 1.60 | 1.58 | 1.59 |
| | n | 95 | 84 | 179 |
| *Total* | M | 3.98 | 3.65 | 3.82 |
| | SD | 1.73 | 1.52 | 1.64 |
| | n | 187 | 173 | 360 |

**Panel D. Morality**

| **Behavior** | | **Proclamation** | | |
|---|---|---|---|---|
| | | Absolute | Control | *Total* |
| Selfish Lie | M | 3.25 | 2.80 | 3.03 |
| | SD | 1.55 | 1.34 | 1.47 |
| | n | 92 | 89 | 181 |
| Prosocial Lie | M | 4.60 | 4.18 | 4.40 |
| | SD | 1.27 | 1.35 | 1.32 |
| | n | 95 | 84 | 179 |
| *Total* | M | 3.94 | 3.47 | 3.71 |
| | SD | 1.56 | 1.51 | 1.55 |
| | n | 187 | 173 | 360 |

**Panel E. Guilt**

| **Behavior** | | **Proclamation** | | |
|---|---|---|---|---|
| | | Absolute | Control | *Total* |
| Selfish Lie | M | 3.38 | 3.13 | 3.26 |
| | SD | 1.48 | 1.49 | 1.48 |
| | n | 92 | 89 | 181 |
| Prosocial Lie | M | 3.89 | 3.27 | 3.60 |
| | SD | 1.41 | 1.57 | 1.52 |
| | n | 95 | 84 | 179 |
| *Total* | M | 3.64 | 3.20 | 3.43 |
| | SD | 1.46 | 1.52 | 1.51 |
| | n | 187 | 173 | 360 |

## 2.2 Supplemental Study 2

### Abbreviated Methods

As in Supplemental Study 1, communicators in Supplemental Study 2 were past study participants who had completed a survey on beliefs and preferences and played a deception game with another person in the previous study. In Supplemental Study 2, we compared evaluations of communicators who endorsed the absolute versus flexible proclamation; communicators told either a prosocial or selfish lie in the deception game.

We intended to recruit 400 participants from Prolific. Participants had to answer comprehension check questions correctly throughout the survey to progress to the end of the study. Of the 367 participants who completed the survey, 40 responses were eliminated due to repeated use of a Prolific ID or failure to provide consent, resulting in a final sample of 327 participants ($N$ = 327, 45.26% female, $M_{age}$ = 30.16, $SD_{age}$ = 10.28). Participants were randomly assigned to one of four conditions in a 2 (proclamation: absolute vs. flexible) x 2 (behavior: prosocial lie vs. selfish lie) between-subjects design.

As in Supplemental Study 1, participants rated communicators on perceived morality, hypocrisy, benevolence, and guilt, after learning about the communicator's survey responses and behavior. We also added a different measure of attitudinal trustworthiness (distinct from The Choice Game). Lastly, participants were asked to provide basic demographic information and explain the intended purpose of The Coin Flip Game before payment and exiting.

***Dependent Variables***

**Morality**. Participants completed the same morality scale as in Supplemental Study 1 with the addition of one new item: "ethical." These items were combined due to high reliability ($r(325) = .91$, $p < .001$).

**Hypocrisy.** Participants indicated agreement to the same statement measured in Supplemental Study 1, as well as two new items: "This person behaved inconsistently with his/her values" and "This person's survey responses conflicted with his/her actions" (1 = Strongly disagree to 7 = Strongly agree). These three items were combined into one composite measure due to high reliability ($\alpha = .86$).

**Guilt**. The same scale was used as in Supplemental Study 1.

**Benevolence**. Benevolence was measured by asking participants to rate communicators on four related traits: "Kind," "Benevolent," "Honest," and "Selfish" [reverse-coded] (1 = Not at all to 7 = Extremely). First, we examined each item independently, since past work suggests that people distinguish benevolence (good intentions) and honesty (making truthful statements), (i.e., Levine & Schweitzer, 2014), however, we found that these items loaded together in our studies, and thus, consistent with our preregistration, we combined them into one composite measure. Items were combined due to high reliability ($\alpha = .88$).

**Attitudinal Trust.** Attitudinal trust was assessed by asking participants to indicate the extent to which the communicator was "Trustworthy" (1 = Not at all to 7 = Extremely).

**Results**

A series of ANOVAs were conducted to examine the effects of proclamation (absolute vs. flexible) and behavior (prosocial lie vs. selfish lie) on each attitudinal variable. Significant interactions were followed with planned simple effect tests (two-sample t-tests), per our pre-registration. Descriptive statistics for Supplemental Study 2 are presented in Table SI-19.

***Morality***

There was a main effect of behavior on moral evaluations ($F(1, 323) = 160.43$, $p < .001$, $\eta_p^2 = .33$). Communicators who told prosocial lies were perceived as more moral than communicators who told selfish lies. The main effect of proclamation and the proclamation x behavior interaction were not significant at the $p < .05$ level ($ps > .796$). Despite being seen as more hypocritical, communicators who endorsed the absolute proclamation and then lied were not viewed as less moral than communicators who endorsed the flexible proclamation before lying.

***Hypocrisy***

There was a main effect of proclamation on hypocrisy ($F(1, 323) = 91.04$, $p < .001$, $\eta_p^2 = .22$), such that communicators who endorsed the absolute proclamation were seen as more hypocritical than communicators who endorsed the flexible proclamation. There was also a main effect of behavior ($F(1, 323) = 27.16$, $p < .001$, $\eta_p^2 = .08$). Communicators who told selfish lies were viewed as more hypocritical than communicators who told prosocial lies. These effects were further qualified by a proclamation x behavior interaction ($F(1, 323) = 5.92$, $p = .015$, $\eta_p^2 = .02$). The effect of proclamation on hypocrisy ratings – namely that the absolute proclamation increases perceptions of hypocrisy relative to the flexible proclamation – was stronger when the subsequent lie was selfish ($t(164) = 8.76$, $p < .001$, $d = 1.36$) compared to when it was prosocial ($t(159) = 4.86$, $p < .001$, $d = 0.77$).

***Guilt***

Neither main effect of proclamation, behavior, nor the proclamation x behavior interaction were significant at the $p < .05$ level ($ps > .075$).

***Attitudinal Trust***

There was a main effect of behavior on attitudinal trust ($F(1, 323) = 126.86$, $p < .001$, $\eta_p^2 = .28$). Communicators who told prosocial lies were seen as more trustworthy than communicators who told selfish lies. The main effect of proclamation and the proclamation x behavior interaction were not significant at the $p < .05$ level ($ps > .791$).

***Benevolence***

There was a main effect of behavior on benevolence ($F(1, 323) = 195.13$, $p < .001$, $\eta_p^2 = .38$). Communicators who told prosocial lies were viewed as more benevolent than communicators who told selfish lies. Neither the main effect of proclamation nor the proclamation x behavior interaction were significant at the $p < .05$ level ($ps > .730$).

**Table SI-19**

*Descriptive Statistics in Supplemental Study 2*

**Panel A. Hypocrisy**

| **Behavior** | | **Proclamation** | | |
|---|---|---|---|---|
| | | Absolute | Flexible | *Total* |
| Selfish Lie | M | 5.19 | 3.35 | 4.27 |
| | SD | 1.55 | 1.12 | 1.64 |
| | n | 83 | 83 | 166 |
| Prosocial Lie | M | 4.01 | 2.92 | 3.49 |
| | SD | 1.61 | 1.20 | 1.52 |
| | n | 84 | 77 | 161 |
| *Total* | M | 4.60 | 3.14 | 3.88 |
| | SD | 1.68 | 1.17 | 1.63 |
| | n | 167 | 160 | 327 |

**Panel B. Morality**

| **Behavior** | | **Proclamation** | | |
|---|---|---|---|---|
| | | Absolute | Flexible | *Total* |
| Selfish Lie | M | 2.63 | 2.69 | 2.66 |
| | SD | 1.37 | 1.25 | 1.31 |
| | n | 83 | 83 | 166 |
| Prosocial Lie | M | 4.55 | 4.56 | 4.56 |
| | SD | 1.40 | 1.40 | 1.40 |
| | n | 84 | 77 | 161 |
| *Total* | M | 3.59 | 3.59 | 3.59 |
| | SD | 1.68 | 1.62 | 1.65 |
| | n | 167 | 160 | 327 |

**Panel C. Guilt**

| **Behavior** | | **Proclamation** | | |
|---|---|---|---|---|
| | | Absolute | Flexible | *Total* |
| Selfish Lie | M | 3.28 | 3.02 | 3.15 |
| | SD | 1.52 | 1.51 | 1.52 |
| | n | 83 | 83 | 166 |
| Prosocial Lie | M | 3.63 | 3.27 | 3.46 |
| | SD | 1.70 | 1.42 | 1.58 |
| | n | 84 | 77 | 161 |

| **Panel C. Guilt (Cont'd)** | | | | |
|---|---|---|---|---|
| | M | 3.46 | 3.14 | 3.30 |
| *Total* | SD | 1.62 | 1.47 | 1.55 |
| | n | 167 | 160 | 327 |

| **Panel D. Attitudinal Trust** | | | | |
|---|---|---|---|---|
| **Behavior** | | **Proclamation** | | |
| | | Absolute | Flexible | *Total* |
| | M | 2.31 | 2.34 | 2.33 |
| Selfish Lie | SD | 1.47 | 1.18 | 1.33 |
| | n | 83 | 83 | 166 |
| | M | 4.10 | 4.16 | 4.12 |
| Prosocial Lie | SD | 1.56 | 1.54 | 1.54 |
| | n | 84 | 77 | 161 |
| | M | 3.21 | 3.21 | 3.21 |
| *Total* | SD | 1.76 | 1.64 | 1.70 |
| | n | 167 | 160 | 327 |

| **Panel E. Benevolence** | | | | |
|---|---|---|---|---|
| **Behavior** | | **Proclamation** | | |
| | | Absolute | Flexible | *Total* |
| | M | 3.57 | 3.45 | 3.51 |
| Selfish Lie | SD | 1.64 | 1.42 | 1.53 |
| | n | 83 | 83 | 166 |
| | M | 6.02 | 6.01 | 6.01 |
| Prosocial Lie | SD | 1.70 | 1.70 | 1.69 |
| | n | 84 | 77 | 161 |
| | M | 4.80 | 4.68 | 4.74 |
| *Total* | SD | 2.07 | 2.01 | 2.04 |
| | n | 167 | 160 | 327 |

***Mediation***

A multiple mediation analysis was conducted to further investigate the effect of proclamation on moral evaluations. Proclamation was entered as the independent variable (0 = absolute, 1 = flexible), ratings of hypocrisy and guilt as simultaneous mediators, and moral evaluations as the dependent variable. We used a bootstrap procedure in the lavaan package in R

(Rosseel, 2012) and found an indirect effect through hypocrisy (Indirect Effect through Hypocrisy = 0.644, 95% CI = [0.44, 0.88]), but no indirect effect through guilt (Indirect Effect through Guilt = -0.077, 95% CI = [-0.19, 0.01]) on moral evaluations.

## 2.3 Supplemental Study 3

### Abbreviated Methods

In Supplemental Study 3, we investigated whether the benefits of the absolute honesty proclamation would persist when the communicator told repeated lies. We examined evaluations of communicators who made absolute or flexible honesty proclamations and told a single lie or repeated lies by randomly assigning participants to one of four conditions in a 2 (proclamation: absolute vs. flexible) x 2 (lie amount: single lie vs. repeated lies) between-subjects design.

We intended to recruit 600 participants from Prolific. Participants had to answer comprehension check questions correctly throughout the survey to progress to the end of the study. 590 participants fully completed the study and there were no further exclusions ($N$ = 590, 50.5% female, $M_{age}$ = 34.13, $SD_{age}$ = 11.5). Participants were introduced to a communicator who endorsed an absolute or flexible honesty proclamation in a survey about beliefs and preferences during a previous study.

In the same previous study, in addition to answering the survey about beliefs and preferences, communicators had to choose whether to lie or tell the truth about a coin flip outcome one time (single lie) or five times (repeated lies) in a game called the Random Outcome Task. In this game, communicators reported the outcome of a single coin flip or a series of coin flips, depending on assigned condition, and they were incentivized to lie when the coin landed on tails in the Random Outcome Task. Participants had to answer comprehension questions about the Random Outcome Task to progress in the survey.

First, participants just learned about the communicator's (absolute or flexible) honesty stance. Participants evaluated the communicators in regard to perceived lying behavior and honesty ideals.

Next, participants learned about the communicator's behavior in the Random Outcome Task. Participants evaluated the communicator a second time in terms of perceived lying behavior and honesty ideals. Participants also evaluated the communicator in terms of morality and hypocrisy just at T2. We asked participants an additional question to assess how much they thought the communicators viewed their own behavior in the Random Outcome Task as lying. Finally, participants answered basic demographic questions before payment and exiting.

***Dependent Variables***

**Morality.** Participants rated the communicator in regard to the traits, "ethical," "moral," and "trustworthy" (1 = Strongly disagree to 7 = Strongly agree). These three items were combined into one composite measure due to high reliability ($\alpha$ =.94).

**Hypocrisy.** Participants rated the communicator in regard to the following questions: "This person is a hypocrite," "This person behaved inconsistently with his/her values"; "This person is inauthentic," "This person is disingenuous." (1 = Strongly disagree to 7 = Strongly agree). These four items were combined into one composite measure due to high reliability ($\alpha$ = .86).

**Lying Behavior.** Participants answered the following question about the communicator to assess perceptions of lying behavior: "How often do you think this person **actually lies?"** (1= Never, 2 = Rarely, 3 = Sometimes, 4 = Often, 5 = Always; bold text appeared in actual survey).

**Honesty Ideals**. Participants answered the following question about the communicator to assess perceptions of honesty ideals: "How often do you think this person **actually thinks it is**

**okay to lie?"** (1= Never, 2 = Rarely, 3 = Sometimes, 4 = Often, 5 = Always; bold text appeared in actual survey)

**Count as Lying.** We were interested to see whether participants thought that the communicator believed the lie(s) in the Random Outcome Task "counted" as a lie. We asked participants to indicate the extent to which they thought, "This person believes that his/her behavior during the Random Outcome Task counts as lying," (1= Strongly disagree to 7 = Strongly agree) in relation to the communicator.

**Results**

***Between-subjects variables (hypocrisy, morality)***

A series of ANOVAs were conducted to examine the effects of proclamation (absolute vs. flexible) and lie amount (single lie vs. repeated lies) on hypocrisy and morality at T2. See Table SI-20 for descriptive statistics.

**Morality.** There was a main effect of lie amount on moral evaluations ($F(1, 586) = 7.57$, $p = .006$, $\eta_p^2 = .01$), such that communicators who told repeated lies ($M = 2.65$, $SD = 1.23$) were seen as less moral than communicators who told a single lie ($M = 2.92$, $SD = 1.17$). The main effect of proclamation and proclamation x lie amount interaction were not significant at the $p < .05$ level ($ps > .583$).

**Hypocrisy.** There was a main effect of proclamation on hypocrisy ($F(1, 586) = 195.23$, $p < .001$, $\eta_p^2 = .25$), such that communicators who endorsed the absolute proclamation were seen as more hypocritical ($M = 5.41$, $SD = 1.26$) than communicators who endorsed the flexible proclamation ($M = 4.04$, $SD = 1.12$). Likewise, there was a significant main effect of lie amount ($F(1, 586) = 11.37$, $p < .001$, $\eta_p^2 = .02$), such that communicators who told a single lie ($M = 4.57$, $SD = 1.31$) were seen as less hypocritical than those who told repeated lies ($M = 4.90$, $SD =$

1.41). The proclamation x lie amount interaction was not significant at the $p < .05$ level ($p = .472$).

***Within-subject variables (lying behavior, honesty ideals)***

A series of mixed within-between subjects ANOVAs were conducted to examine the effects of proclamation, lie amount, and time (within-subjects) on perceived lying behavior and honesty ideals.

**Lying Behavior**. The response options of "never," "rarely," "sometimes," "often," and "always," which aligned with the Attitudes Towards Honesty measures in Studies 4-5, were coded from 1 to 5; and therefore, higher numbers indicate greater perceived lying behavior.

Results revealed a main effect of proclamation ($F(1, 586) = 114.95$, $p < .001$, $\eta_p^2 = .16$), such that communicators who endorsed the absolute proclamation ($M = 2.77$, $SD = 0.88$) were thought to lie less frequently than communicators who endorsed the flexible proclamation ($M = 3.26$, $SD = 0.64$). There was also a main effect of lie amount ($F(1, 586) = 4.06$, $p = .044$, $\eta_p^2 < .01$), such that communicators who told a single lie ($M = 2.96$, $SD = 0.76$) were thought to lie less frequently than communicators who told repeated lies ($M = 3.06$, $SD = 0.85$). Additionally, there was a main effect of time ($F(1, 586) = 501.06$, $p < .001$, $\eta_p^2 = .46$); when communicators were evaluated at T2 ($M = 3.36$, $SD = 0.70$) they were thought to lie more frequently than when they were evaluated at T1 ($M = 2.66$, $SD = 0.75$).

These effects were further qualified by a significant proclamation x time interaction ($F(1, 586) = 65.40$, $p < .001$, $\eta_p^2 = .10$). Perceived lying behavior attributed to communicators was greater at T2 than T1 across proclamation conditions (i.e., after learning about the communicator's deception, participants thought communicators were more likely to lie than they originally indicated, regardless of the communicator's proclamation); however, the magnitude of

this difference across time was greater for the communicators who endorsed the absolute proclamation compared to those who endorsed the flexible proclamation. Even though participants adjusted their expectations of perceived lying behavior after learning about deception for all communicators —and adjustments were more substantial for communicators endorsing absolute honesty— communicators who endorsed the absolute proclamation were *still* thought to lie less frequently than communicators who endorsed the flexible proclamation at T2.

There was also a significant lie amount x time interaction ($F(1, 586) = 17.30$, $p < .001$, $\eta_p^2 = .03$), such that all communicators were suspected of lying more frequently when evaluated at T2 than T1, but the difference in perceived lying behavior ratings across time was larger for the repeat liars compared to the single liars.

Neither the proclamation x lie amount interaction nor the three-way proclamation x lie amount x time interaction were significant at the $p < .05$ level (*ps* > .493).

**Honesty Ideals**. The response options of "never," "rarely," "sometimes," "often," and "always," were coded from 1 to 5, such that higher numbers represent more modest honesty ideals and lower numbers represent more ambitious honesty ideals.

There was a main effect of proclamation on honesty ideals ($F(1, 586) = 353.15$, $p < .001$, $\eta_p^2 = .38$). Communicators who endorsed the absolute proclamation were seen as having more ambitious honesty ideals ($M = 2.33$, $SD = 1.12$) than communicators who endorsed the flexible proclamation ($M = 3.24$, $SD = 0.62$). There was a main effect of lie amount ($F(1, 586) = 8.38$, $p = .004$, $\eta_p^2 = .01$), such that communicators who told a single lie ($M = 2.71$, $SD = 0.98$) were seen as having more ambitious honesty ideals than communicators who told repeated lies ($M = 2.85$, $SD = 1.05$). There was also a significant main effect of time ($F(1, 586) = 753.39$, $p < .001$, $\eta_p^2 =$

.56), such that communicators were seen as having more modest honesty ideals at T2 ($M$ = 3.26, $SD$ = 0.83) than T1 ($M$ = 2.30, $SD$ = 0.96).

These effects were further qualified by a significant proclamation x time interaction ($F(1, 586) = 223.89$, $p < .001$, $\eta_p^2 = .28$). Both communicators endorsing absolute and flexible proclamations were seen as having more modest honesty ideals at T2 than T1, but the magnitude of this difference in ratings across time was more extreme for the communicators who endorsed the absolute proclamation compared to the communicators who endorsed the flexible proclamation. Despite this adjustment across time, communicators who endorsed the absolute proclamation were still thought to have more ambitious honesty ideals overall, similarly to the results in the perceived lying behavior model. In other words, participants thought that communicators who endorsed the absolute proclamation both tried to lie less frequently and succeeded at lying less frequently compared to communicators who endorsed the flexible proclamation at both points in time. However, the difference in evaluations of perceived lying behavior and honesty ideals of communicators based on proclamation were greater when participants only knew about the communicator's proclamation and had not yet encountered their inconsistent behavior.

Likewise, there was a significant lie amount x time interaction ($F(1, 586) = 12.97$, $p < .001$, $\eta_p^2 = .02$). The difference in ratings of honesty ideals between communicators who told a single lie and repeated lies was greater at T2 than T1.

Neither the proclamation x lie amount interaction nor the three-way proclamation x lie amount x time interaction were significant at the $p < .05$ level ($ps > .557$).

**Table SI-20**

*Descriptive Statistics in Supplemental Study 3*

| **Panel A. Hypocrisy** | | | | |
|---|---|---|---|---|
| **Behavior** | | **Proclamation** | | |
| | | Absolute | Flexible | *Total* |
| Single Lie | M | 5.28 | 3.85 | 4.57 |
| | SD | 1.18 | 1.02 | 1.31 |
| | n | 150 | 146 | 296 |
| Repeated Lies | M | 5.54 | 4.24 | 4.90 |
| | SD | 1.33 | 1.18 | 1.41 |
| | n | 149 | 145 | 294 |
| *Total* | M | 5.41 | 4.04 | 4.74 |
| | SD | 1.26 | 1.12 | 1.37 |
| | n | 299 | 291 | 590 |
| **Panel B. Morality** | | | | |
| **Behavior** | | **Proclamation** | | |
| | | Absolute | Flexible | *Total* |
| Single Lie | M | 2.87 | 2.97 | 2.92 |
| | SD | 1.21 | 1.14 | 1.17 |
| | n | 150 | 146 | 296 |
| Repeated Lies | M | 2.65 | 2.64 | 2.65 |
| | SD | 1.30 | 1.15 | 1.23 |
| | n | 149 | 145 | 294 |
| *Total* | M | 2.76 | 2.80 | 2.78 |
| | SD | 1.26 | 1.16 | 1.21 |
| | n | 299 | 291 | 590 |
| **Panel C. Perceived Lying Frequency** | | | | |
| **Time** | | **Proclamation** | | |
| **Time One (T1)** | | Absolute | Flexible | *Total* |
| Single Lie | M | 2.30 | 3.06 | 2.68 |
| | SD | 0.73 | 0.59 | 0.77 |
| | n | 150 | 146 | 296 |
| Repeated Lies | M | 2.28 | 3.00 | 2.63 |
| | SD | 0.73 | 0.55 | 0.74 |
| | n | 149 | 145 | 294 |
| Total | M | 2.29 | 3.03 | 2.66 |
| | SD | 0.73 | 0.57 | 0.75 |
| | n | 299 | 291 | 590 |

| **Time Two (T2)** | | Absolute | Flexible | *Total* |
|---|---|---|---|---|
| | M | 3.12 | 3.39 | 3.25 |
| Single Lie | SD | 0.67 | 0.58 | 0.64 |
| | n | 150 | 146 | 296 |
| | M | 3.38 | 3.57 | 3.48 |
| Repeated Lies | SD | 0.81 | 0.64 | 0.74 |
| | n | 149 | 145 | 294 |
| | M | 3.25 | 3.48 | 3.36 |
| *Total* | SD | 0.76 | 0.62 | 0.70 |
| | n | 299 | 291 | 590 |

**Panel B. Honesty Ideals**

| **Time** | | **Proclamation** | | |
|---|---|---|---|---|
| **Time One (T1)** | | Absolute | Flexible | *Total* |
| | M | 1.59 | 3.02 | 2.29 |
| Single Lie | SD | 0.76 | 0.48 | 0.96 |
| | n | 150 | 146 | 296 |
| | M | 1.60 | 3.03 | 2.31 |
| Repeated Lies | SD | 0.73 | 0.55 | 0.96 |
| | n | 149 | 145 | 294 |
| | M | 1.60 | 3.03 | 2.30 |
| *Total* | SD | 0.75 | 0.51 | 0.96 |
| | n | 299 | 291 | 590 |
| **Time Two (T2)** | | Absolute | Flexible | *Total* |
| Single Lie | M | 2.91 | 3.35 | 3.13 |
| | SD | 0.90 | 0.62 | 0.80 |
| | n | 150 | 146 | 296 |
| Repeated Lies | M | 3.22 | 3.57 | 3.39 |
| | SD | 0.96 | 0.65 | 0.84 |
| | n | 149 | 145 | 294 |
| *Total* | M | 3.07 | 3.46 | 3.26 |
| | SD | 0.94 | 0.64 | 0.83 |
| | n | 299 | 291 | 590 |

*Notes.* Morality and hypocrisy were measured only at T2. Perceived lying behavior and honesty ideals were measured at both time points and coded such that lower numbers indicated less perceived lying and more ambitious honesty ideals (1 = never, 2 = rarely, 3 = sometimes, 4 = often, 5 = always, both scales).

***Exploratory Measure***

We conducted a between-subjects ANOVA to examine the effects of proclamation (absolute vs. flexible) and lie amount (single lie vs. repeated lies) on an exploratory "counts as lying" measure. We were interested to see whether participants thought that communicators viewed their own lies in the Random Outcome Task as deception, despite the fact that these lies were minor and took place within a game context.

There was a main effect of proclamation ($F(1, 586) = 6.96$, $p = .009$, $\eta_p^2 = .01$), such that participants believed that communicators who endorsed the absolute proclamation ($M = 4.50$, $SD = 1.68$) were less likely to view their behavior as actual lies relative to communicators who endorsed the flexible proclamation ($M = 4.86$, $SD = 1.64$). There was also a main effect of lie amount ($F(1, 586) = 14.39$, $p < .001$, $\eta_p^2 = .02$). Interestingly, participants thought that communicators viewed a single lie ($M = 4.93$, $SD = 1.60$) as more of a true lie than repeated lies ($M = 4.42$, $SD = 1.70$). The proclamation x lie amount interaction was not significant at the $p < .05$ level ($p = .179$).

***Multiple Mediation Analysis***

We conducted a bootstrapped multiple mediation analysis within the lavaan package in R (Rosseel, 2012) to examine the relationship between the proclamation and moral evaluations and potential mechanisms of construal of lies, lying behavior, and honesty ideals. Proclamation (1 = flexible, 0 = absolute) was entered as the independent variable and moral evaluations as the dependent variable. Our "counts as lying" measure, lying behavior, and honesty ideals were entered as simultaneous mediators. We focus only on T2 judgments of lying behavior and honesty ideals, since participants had learned about the communicator's proclamation and deceptive behavior by T2. Counts as lying was only measured at T2.

Both lying behavior (Indirect Effect through Lying Behavior = -0.183, 95% CI = [-0.29, -0.09]) and honesty ideals (Indirect Effect through Honesty Ideals = -0.112, 95% CI = [-0 .18, -0.05]) significantly mediated the relationship between proclamation and moral evaluations. Counts as lying did not significantly mediate the relationship between proclamation and moral evaluations (Indirect Effect through Counts as Lying = -0.015, 95% CI = [-0.04, 0 .00]). These results indicate that our effects are not due to differences in perceptions of construal of behavior between the communicators endorsing absolute and flexible honesty, while also adding support to our future honesty mechanism, Inferences about lying behavior and honesty ideals are reflected in our future honesty measure (the main mechanism examined in our manuscript), which includes items both about the communicator's honest behavior, and their underlying commitment to honesty.

## 2.4 Supplemental Study 4

### Abbreviated Methods

In Supplemental Study 4, we investigated whether the benefits of the absolute proclamation would remain if the candidate was advised to take an absolute stance on honesty by a political advisor. In other words, the absolute stance could be interpreted as strategic instead of reflecting private beliefs. This study was nearly identical to Studies 1 and 2 of the main manuscript, but we included an advisor condition, in which a political advisor prepped Matthew Johnson to take the absolute or flexible honesty stance during the interview. We examined evaluations of candidates who made absolute or flexible proclamations after being advised or not advised to take the stance by a campaign advisor, and we included one control condition, in which Matthew Johnson did not take any honesty stance (he was not asked about ethics and politics in this condition). Participants were randomly assigned to one of five conditions in a 2

(proclamation: absolute vs. flexible) x 2 (advisor: no advisor vs. political advisor) + control between-subjects design.

We aimed to recruit 500 participants on MTurk. We included attention checks at the end of the survey and excluded participants who did not correctly answer the attention check questions from the final analysis, resulting in 436 participants ($N$ = 436, 46.79% female, $M_{age}$ = 39.15, $SD_{age}$ = 11.12). Participants were introduced to a political candidate named Matthew Johnson who took an absolute or flexible stance on honesty, or no stance on honesty, during a television interview. If Matthew Johnson was described as taking an absolute or flexible stance on honesty, he was either described as making the proclamation with no preparation, as in Studies 1-2, or as being prepped by a campaign advisor before making the proclamation during the interview.

After the television interview, Matthew Johnson was described as engaging in deception about his campaign finances. Then, participants indicated their intentions of voting for Mr. Johnson and evaluated Mr. Johnson in regard to morality, hypocrisy, perceived lying behavior, and honesty ideals. We also included an exploratory measure about the perceived immorality of lying itself.

***Dependent Variables***

**Voting Intentions, Morality, and Hypocrisy.** Participants responded to these measures using the same items as we used in Study 1 of the main manuscript (αs > .89).

**Lying Behavior and Honesty Ideals**. Participants indicated how often they thought Mr. Johnson **actually lies** and **actually thinks it is okay to lie** using similar items to those described in Study S3.

**Perceived Immorality.** We were interested to see how immoral participants thought lying was themselves. We included an additional question asking participants, "Regardless of the scenario that you read, how immoral was it for Matthew Johnson to lie?" (1 = Not at all to 7 = Extremely).

**Results**

A series of ANOVAs were conducted to examine the effects of proclamation (absolute vs. flexible) and advisor prep (no advisor vs. political advisor) on voting intentions and attitudinal evaluations. We excluded the control condition from these 2 x 2 ANOVAs, as was pre-registered.

***Voting Intentions***

There was a main effect of proclamation on voting intentions ($F(1, 346) = 11.34$, $p < .001$, $\eta_p^2 = .03$), such that participants were more likely to vote for the candidate who took the absolute honesty stance ($M = 3.05$, $SD = 1.49$) than the candidate who took the flexible honesty stance ($M = 2.51$, $SD = 1.42$). The main effect of the advisor condition and proclamation x advisor interaction were not significant at the $p < .05$ level ($ps > .330$).

***Morality***

There was a main effect of proclamation on moral evaluations ($F(1, 346) = 7.11$, $p = .008$, $\eta_p^2 = .02$), such that the candidate who took the absolute honesty stance ($M = 2.81$, $SD = 1.32$) was seen as more moral than the candidate who took the flexible honesty stance ($M = 2.44$, $SD = 1.28$). The main effect of the advisor condition and proclamation x advisor interaction were not significant at the $p < .05$ level ($ps > .589$).

***Hypocrisy***

Results revealed a main effect of proclamation on hypocrisy ($F(1, 346) = 9.73$, $p = .002$, $\eta_p^2 = .03$), such that the candidate who took the absolute honesty stance ($M = 5.45$, $SD = 1.28$) was seen as more hypocritical than the candidate who took the flexible honesty stance ($M = 5.02$, $SD = 1.30$). The main effect of the advisor condition and proclamation x advisor interaction were not significant at the $p < .05$ level ($ps > .085$).

***Lying Behavior***

There was a main effect of proclamation on lying behavior ($F(1, 346) = 9.64$, $p = .002$, $\eta_p^2 = .03$), such that the candidate who made the flexible proclamation about honesty was seen as actually lying more ($M = 4.95$, $SD = 1.17$) than the candidate who made the absolute proclamation about honesty ($M = 4.55$, $SD = 1.20$).

This was further qualified by a significant proclamation x advisor interaction on perceptions of lying behavior for Matthew Johnson ($F(1, 346) = 9.60$, $p = .002$, $\eta_p^2 = .03$). When Matthew Johnson was advised to take the absolute honesty stance, participants thought he actually lied more frequently ($M = 4.67$, $SD = 1.21$) compared to when Matthew Johnson took the absolute honesty stance with no advisor input ($M = 4.42$, $SD = 1.19$). Conversely, when Matthew Johnson was advised to take the flexible honesty stance, participants thought he actually lied *less* frequently ($M = 4.67$, $SD = 1.12$) compared to when he took the flexible honesty stance without any advisor input ($M = 5.20$, $SD = 1.16$).

The main effect of the advisor condition was not significant at the $p < .05$ level ($p = .273$).

***Honesty Ideals***

There was a main effect of proclamation on honesty ideals ($F(1, 346) = 18.11$, $p < .001$, $\eta_p^2 = .05$), such that the candidate who took the flexible honesty stance was seen as having more

modest honesty ideals ($M$ = 5.14, $SD$ = 1.38) than the candidate who took the absolute honesty stance ($M$ = 4.51, $SD$ = 1.43).

This was further qualified by a significant proclamation x advisor interaction ($F(1, 346)$ = 11.34, $p < .001$, $\eta_p^2 = .03$). Similar to the lying behavior model, when Matthew Johnson was advised to take the absolute honesty stance, his honesty ideals seemed more modest ($M$ = 4.77, $SD$ = 1.34) compared to when he took the absolute honesty stance with no explicit advisor input ($M$ = 4.24, $SD$ = 1.48). The reverse effect occurred for the candidate who took the flexible honesty stance: When Matthew Johnson took the flexible honesty stance *without* any advisor input, participants thought he had more modest honesty ideals ($M$ = 5.37, $SD$ = 1.27) compared to when he took the flexible honesty stance with advisor input ($M$ = 4.90, $SD$ = 1.46).

The main effect of the advisor condition was not significant at the $p < .05$ level ($p$ = .805).

***Control Comparisons***

We conducted a series of two-sample t-tests to compare the control candidate versus the candidate endorsing absolute honesty with no advisor preparation versus the candidate endorsing flexible honesty with no advisor preparation. When comparing moral evaluations of the candidates in the absolute and flexible proclamation conditions (with no advisor prep), we find that the candidate who endorsed absolute honesty is viewed as more moral ($t(173)$ = 2.35, $p$ = .020, $d$ = 0.35) than the candidate who endorsed flexible honesty. Additionally, the candidate in the control condition is viewed as more moral than the candidate who endorsed flexible honesty ($t(174)$ = 2.39, $p$ = .018, $d$ = 0.36). However, the candidate who endorsed absolute honesty and the control candidate were not viewed significantly differently ($t(169)$ = -0.17, $p$ = .868, $d$ = -0.03).

***Perceived Immorality***

We conducted a linear regression to see whether participants' own ratings of the perceived immorality of lying interacted with the proclamation to predict moral evaluations. Results revealed a main effect of perceived immorality ($b$ = -0.19, $p$ = .002), such that as ratings of the immorality of lying went up, moral ratings of the candidate went down. The main effect of proclamation and perceived immorality x proclamation interaction were not significant at the $p <$ .05 level ($ps$ > .181).

## 2.5 Supplemental Study 5

### Abbreviated Methods

In Supplemental Study 5, we compared evaluations of communicators who made absolute or flexible honesty proclamations and told selfish lies. We also began to explore the robustness of our effects across cultures. We asked participants, from both the United States and India, to evaluate communicators at two points in time: once after their stance on honesty was revealed, and a second time after they told a selfish lie (order was *not* randomized). Participants were randomly assigned to a proclamation condition (absolute vs. flexible) and all participants were asked to evaluate communicators twice (T1 vs. T2) in a mixed between-within-subjects design. At the end of the study, we also assessed participants' willingness to take advice from the communicator using a modified Weight of Advice task that we referred to as the Advice Game (Gino & Schweitzer, 2008).

We had three goals in Supplemental Study 5: 1) to replicate our findings on the initial costs of moral flexibility in Studies 2-3, 2) to consider the influence of hypocritical absolutism on willingness to take advice, and 3) to explore potential cultural moderation.

The stimuli and design of this study were very similar to Study 3 reported in the main manuscript. There were, however, three key differences. First, all participants evaluated the communicator in the same order: first after learning about their honesty stance (manipulated through screenshots from a previous survey on beliefs and preferences), and again after learning about the communicator's deception (manipulated by describing the communicator's behavior in the Coin Flip Game). Second, the communicator was always described as telling a *selfish* (vs. prosocial) lie in the Coin Flip Game. Third, Supplemental Study 5 included the Advice Game as another dependent variable.

We aimed to recruit 200 participants from the United States and India on MTurk (100 participants in the US and 100 participants in India). Participants who failed to answer attention check questions correctly were excluded from analysis, resulting in 196 participants ($N$ = 196, 36.22% female, $M_{age}$ = 34.63, $SD_{age}$ = 9.43). Participants were matched with a previous participant who was assigned to the role of the Advisor for the Advice Game. The past participant was described with a random number and referred to as "Advisor [random number]" (e.g., Advisor 86) throughout the study (and hereafter referred to as the Advisor). In part one of the Advice Game, the participant was shown a picture of a jar of coins and asked to make an initial estimate about the amount of money in a jar of coins. Next, the participant learned about how the Advisor had answered a survey about beliefs and preferences in a previous study, which we used to manipulate the Advisor's (absolute or flexible) honesty stance. Participants evaluated the Advisor knowing just this information at T1.

Then, participants learned about the Advisor's selfish lie behavior (again from a previous study) and evaluated the Advisor a second time. Finally, in part two of the Advice Game, the Advisor sent the participant advice about how much money was in the jar of coins. The

participant had the chance to revise their initial estimate before submitting a final estimate as the final step of the Advice Game. Participants knew there was an incentive for the Advisor to lie *or* to tell the truth, but they did not know the details of the incentive (i.e., whether to lie or tell the truth, how much money was involved). Therefore, they had to decide whether to trust the advice of the Advisor without knowing if the Advisor was incentivized to lie. In this way, revising the final estimate to an amount closer to the Advisor's advised amount indicates trust in the Advisor to tell the truth regardless of potential incentives.

### *Dependent Variables*

**Willingness to Take Advice.** Participants had the option to revise their initial estimate in the Advice Game, after receiving advice from the Advisor on the amount of money in the jar of coins. To calculate weight of advice, we divided the absolute value of the difference between the participant's final estimate and initial estimate by the absolute value of the difference between the amount the Advisor said was in the jar of coins (the advice) and the participant's initial estimate. The resulting weight of advice ratio measure reflects the degree to which participants weighted their own initial estimate (0 means that participants' final estimate was the same as their initial estimate, so the advice had no influence on their final estimate) relative to the Advisor's advice (1 means that the participants' final estimate was the same as the advice, so they fully relied on the Advisor's advice). Although the majority of participant responses ranged between 0 (ignore advice) to 1 (fully rely on advice), there were nine participants who submitted final estimates that over relied on the advice, resulting in ratios over 1.

**Morality.** Participants rated the Advisor using the same morality scale as in Studies 3-5 ($\alpha$ = .93, T1; $\alpha$ = .97, T2).

**Hypocrisy**. Participants rated the Advisor using a similar hypocrisy scale as in Studies 3-5 but modified to be three items instead of five items (α = .95, T1; α = .94, T2).

**Future Honesty.** Participants rated the Advisor using the same future honesty scale as in Studies 3-5 (α = .67, T1; α = .70, T2).

**Results**

We conducted a one-way ANOVA to examine the effects of proclamation (absolute vs. flexible) on willingness to take advice. A series of mixed ANOVAs were conducted to examine the effects of proclamation (absolute vs. flexible) and time (T1 vs. T2) on attitudinal evaluations.

***Willingness to Take Advice***

One subject was removed from this analysis because their initial estimate was equal to the amount of money that the Advisor later advised was actually in the jar of coins ($9.24). This subject kept the same final estimate as the initial estimate ($9.24), but it is not possible to confirm whether this is because the subject relied fully on their own intuition or fully on the Advisor's advice, since either situation would lead to the same estimate.

A one-way ANOVA on the remaining 195 subjects, revealed no significant main effect of proclamation ($p$ = .461). Participants did not rely on advice from Advisors who took an absolute stance and told a selfish lie ($M$ = 0.56, $SD$ = 0.58) significantly more than Advisors who took a flexible stance and told a selfish lie ($M$ = 0.51, $SD$ = 0.41). Thus, while hypocritical absolutism did not significantly enhance willingness to take advice relative to consistent flexibility, it also did not detract from willingness to take advice.

***Morality***

There was a main effect of proclamation on moral evaluations ($F(1, 194) = 12.32$, $p < .001$, $\eta_p^2 = .06$), such that Advisors who took the absolute honesty stance were seen as more

moral than Advisors who took the flexible stance. There was also a main effect of time ($F$(1, 194) = 131.52, $p$ < .001, $\eta_p^2$ = .40), such that Advisors were seen as more moral at T1 than T2. These effects were qualified by a significant proclamation x time interaction ($F$(1, 194) = 23.55, $p$ < .001, $\eta_p^2$ = .11). Advisors who endorsed absolute honesty and lied were seen as more moral than Advisors who endorsed flexible honesty and lied overall, but the discrepancy in moral evaluations of these types of Advisors was much larger at T1 ($M_{absolute}$ = 5.77, $SD_{absolute}$ = 0.82; $M_{flexible}$ = 4.47, $SD_{flexible}$ = 1.46) than at T2 ($M_{absolute}$ = 3.68, $SD_{absolute}$ = 2.03; $M_{flexible}$ = 3.63, $SD_{flexible}$ = 1.89).

***Hypocrisy***

There was a main effect of time on hypocrisy evaluations ($F$(1, 194) = 125.40, $p$ < .001, $\eta_p^2$ = .39), such that the Advisors were seen as more hypocritical at T2 than T1. There was also a significant proclamation x time interaction ($F$(1, 194) = 23.66, $p$ < .001, $\eta_p^2$ = .11). Advisors who took the flexible stance were seen as more hypocritical at T1 than advisors who took the absolute honesty stance ($M_{absolute}$= 3.05, $SD_{absolute}$ = 1.80; $M_{flexible}$= 3.93, $SD_{flexible}$ = 1.57), but as less hypocritical at T2 ($M_{absolute}$ = 5.14, $SD_{absolute}$ = 1.45; $M_{flexible}$ = 4.75, $SD_{flexible}$ = 1.52). The main effect of proclamation was not significant $p$ < .05 level ($p$ = .197).

***Future Honesty***

There was a main effect of proclamation on future honesty ($F$(1, 194) = 35.18, $p$ < .001, $\eta_p^2$ = .15), such that Advisors who took the absolute stance were seen as more likely to engage in future honesty than Advisors who took the flexible stance. There was also a main effect of time ($F$(1, 194) = 117.27, $p$ < .001, $\eta_p^2$ = .38), such that Advisors were seen as more likely to engage in future honesty at T1 than T2. These effects were qualified by a significant proclamation x time interaction ($F$(1, 194) = 27.11, $p$ < .001, $\eta_p^2$ = .12). Advisors who endorsed absolute honesty and

lied were seen as more likely to engage in future honesty than Advisors who endorsed flexible honesty and lied overall, but the discrepancy in future honesty evaluations was much larger at T1 ($M_{absolute}$ = 5.15, $SD_{absolute}$ = 1.05; $M_{flexible}$ = 3.86, $SD_{flexible}$ = 0.92) than T2 ($M_{absolute}$ = 3.38, $SD_{absolute}$ = 1.33; $M_{flexible}$ = 3.24, $SD_{flexible}$ = 1.23).

***Country Effects***

To determine whether culture moderated our effects, we conducted the same set of analyses with country (India vs. US) added as a factor. Overall, accounting for country did not reverse or eliminate our effects, but hypocrisy evaluations (and associated reductions in advice-taking, moral evaluations, and future honesty evaluations) were stronger in evaluations from American than Indian participants (consistent with Effron, Markus, et al., 2018). We report the full results from this series of ANOVAs in Table SI-21 and summarize these results briefly below.

**Willingness to Take Advice.** There was a main effect of country ($F(1, 191) = 6.08$, $p = .015$, $\eta_p^2 = .03$), such that advice taking was lower among American participants ($M = 0.45$, $SD = 0.39$) relative to Indian participants ($M = 0.63$, $SD = 0.58$ ). However, there were not a significant main effect of proclamation or proclamation x country interaction ($ps > .217$).

**Attitudinal Variables.** For our attitudinal measures, a main effect of time and interaction between proclamation x time persisted in the models on morality, hypocrisy, and future honesty, as did the main effect of proclamation in the models on morality and future honesty. However, there was also a significant effect of country, country x time, and three-way culture x time x proclamation interaction on evaluations of hypocrisy, morality, and future honesty, as well as a significant proclamation x country interaction on evaluations of future honesty. These results are presented in Table SI-21 and Figure SI-2 and described in more detail below.

Overall, Advisors who took an absolute stance were seen as more moral and more likely to engage in future honesty at T1 relative to Advisors who took a flexible stance in both countries, highlighting the costliness of admitting to moral flexibility at T1 (consistent with Studies 2 and 3 of the main manuscript). When these Advisors then lied, they were seen as hypocritical in both countries, but American participants viewed the Advisors who took an absolute stance as more hypocritical at T2 (American participants: $M_{absolute\ time\ two}$ = 5.55, $SD_{absolute\ time\ two}$ = 1.25) compared to the Indian participants (Indian participants: $M_{absolute\ time\ two}$ = 4.79, $SD_{absolute\ time\ two}$ = 1.53). These hypocrisy evaluations influenced moral and future honesty evaluations. American participants rated the Advisors who took an absolute stance and lied as less moral relative to the Indian participants at T2 (American participants: $M_{absolute\ time\ two}$ = 2.33, $SD_{absolute\ time\ two}$ = 1.31; Indian participants: $M_{absolute\ time\ two}$ = 4.83, $SD_{absolute\ time\ two}$ = 1.83), and as less likely to engage in future honesty (American participants: $M_{absolute\ time\ two}$ = 2.66, $SD_{absolute\ time\ two}$ = 1.18; Indian participants: $M_{absolute\ time\ two}$ = 3.99, $SD_{absolute\ time\ two}$ = 1.13; see Figure SI-2). Importantly, however, Advisors who endorsed flexible honesty and lied were *also* penalized in both countries at T2 (see Figure SI-2). These results suggest that telling a selfish lie was detrimental *overall*, but telling a selfish lie *and* behaving hypocritically did not have significant added costs relative to telling a selfish lie that was consistent with a flexible honesty stance. Advisors who took an absolute stance and lied were not seen as any less moral than the Advisors who took a flexible stance and lied; there were no significant differences in morality and future honesty evaluations for Advisors endorsing absolute or flexible honesty at T2 in either country (morality, US: $t(98) = -1.35$, $p = .180$, $d = -0.27$; India: $t(94) = 0.07$, $p = .945$, $d = 0.01$; future honesty, US: $t(98) = -0.14$, $p = .889$, $d = -0.03$; India: $t(94) = 0.20$, $p = .842$, $d = 0.04$).

These country effects are consistent with past cross-cultural hypocrisy research, adding support to the notion that hypocrisy penalties may be greater in independent than interdependent cultures (Effron, Markus, et al., 2018). However, we are hesitant to draw strong cultural conclusions given that we measured culture by including a participant variable of country, with no additional direct measures of culture, and because we only recruited participants from two countries. Furthermore, we were unable to filter our MTurk recruitment by worker quality when recruiting participants in India, as we do in our other MTurk studies, which may have influenced the quality and reliability of this dataset. Yet, these exploratory findings allude to potential cross-cultural differences in perceptions of hypocrisy that are consistent with previous work.

**Table SI-21**

*ANOVA Results from Supplemental Study 5*

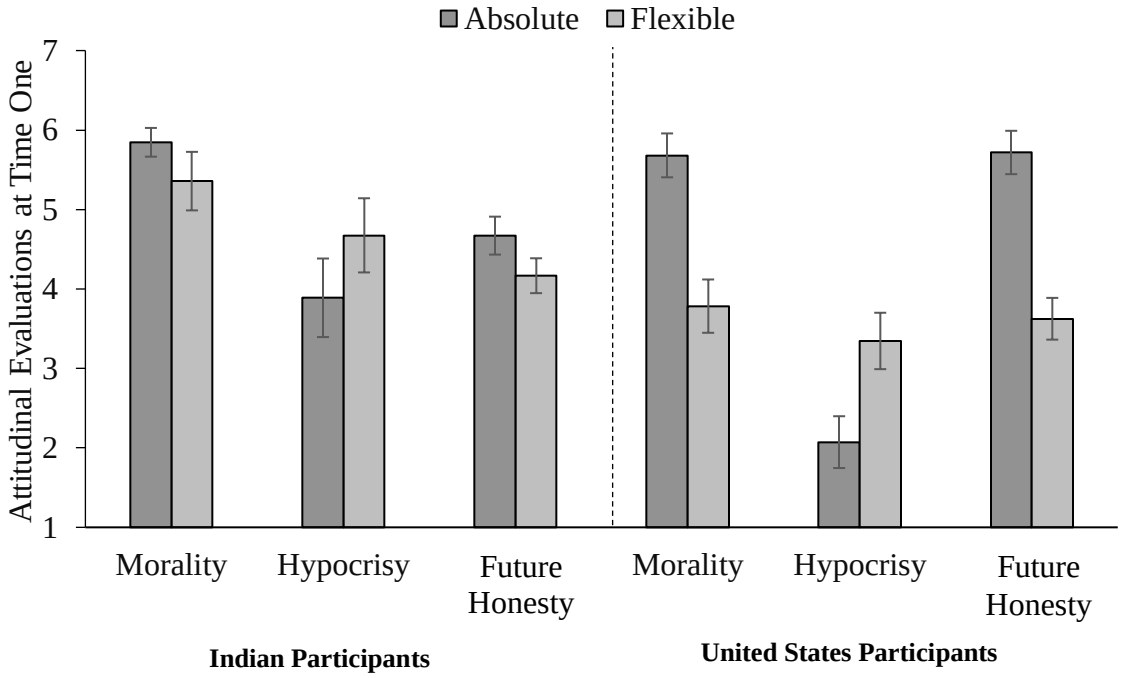
| **Variable** | **Willingness to Take Advice** | | **Morality** | | **Hypocrisy** | | **Future Honesty** | |
|---|---|---|---|---|---|---|---|---|
| | F | *p* | F | *p* | F | *p* | F | *p* |
| Proclamation | 0.26 | .609 | 10.48 | .001** | 2.91 | .090 | 33.48 | < .001*** |
| Time | | | 173.35 | < .001*** | 180.07 | < .001*** | 178.03 | < .001*** |
| Country | 6.08 | .015* | 101.22 | < .001*** | 12.71 | < .001*** | 21.39 | < .001*** |
| Proclamation x Time | | | 36.11 | < .001*** | 42.88 | < .001*** | 49.63 | < .001*** |
| Proclamation x Country | 1.53 | .217 | 2.56 | .111 | 0.45 | .501 | 11.23 | < .001*** |
| Time x Country | | | 39.11 | < .001*** | 70.65 | < .001*** | 71.11 | < .001*** |
| Proclamation x Country x Time | | | 15.81 | < .001*** | 11.18 | < .001*** | 20.57 | < .001*** |

*Notes.* Willingness to take advice was only measured once and this model only included proclamation, country, proclamation x country. *$p$ <.05. ** $p$ < .01. *** $p$ < .001.

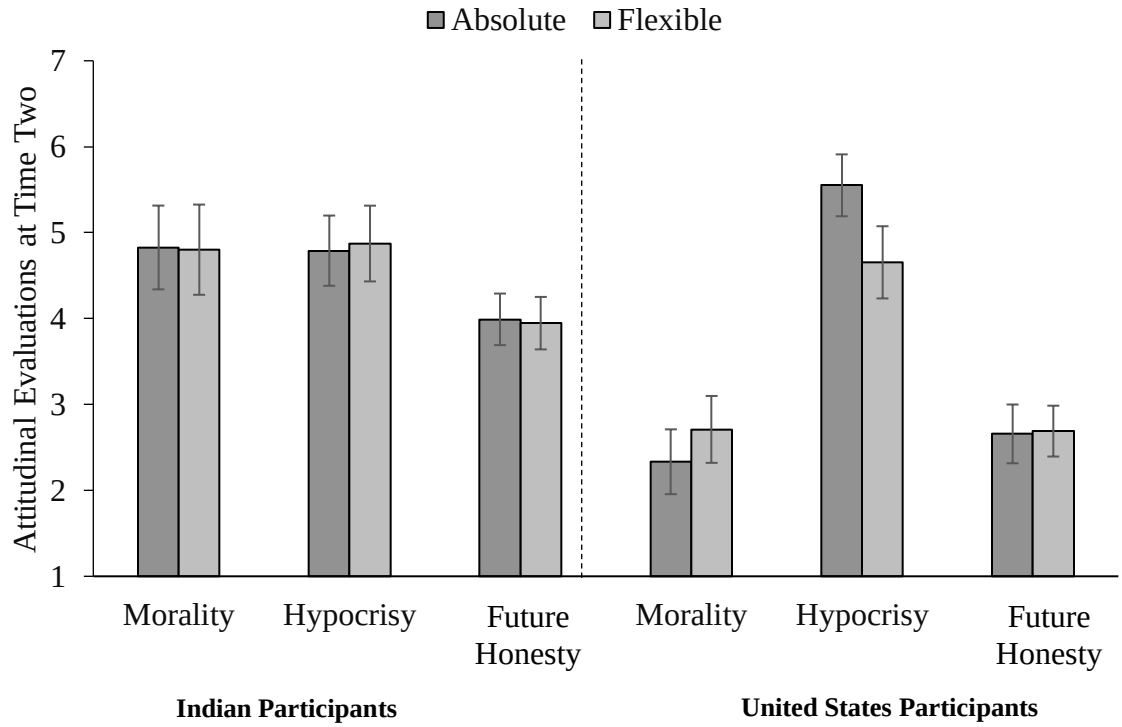
**Figure SI-2**

*Advisor Attitudinal Evaluations in Supplemental Study 5*

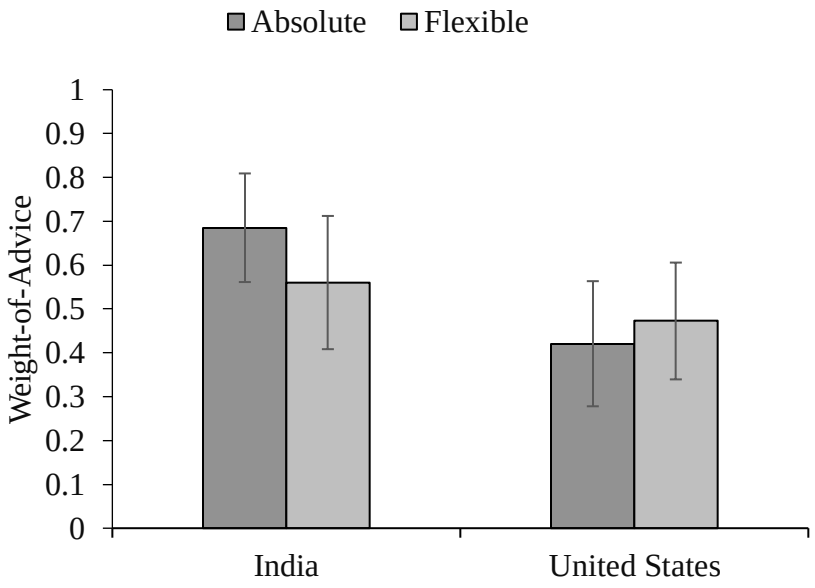
**Panel A. Attitudinal Evaluations at Time One in India and the United States**




**Panel B. Attitudinal Evaluations at Time Two in India and the United States**



**Panel C. Weight of Advice in India and the United States**




*Note.* In this ratio measure, higher scores reflect greater reliance on the Advisor's advice when submitting the final estimate.

## 2.6 Supplemental Study 6

### Abbreviated Methods

In Supplemental Study 6, we investigated how the broadness (vs. specificity) of the absolute honesty proclamation influenced judgments of hypocrisy and morality. We were curious to see if violating a more specific proclamation would seem more disingenuous and self-serving, and thus lead to greater perceptions of hypocrisy and lower perceptions of morality.

We conducted a study in the political domain that was similar to Study 2. However, we manipulated the specificity of the absolute honesty stance during the interview between Matthew Johnson and a television news anchor (we did not examine the flexible honesty proclamation in this study). Participants were randomly assigned to one condition in a between-subjects design (broadness: broad absolute proclamation vs. specific absolute proclamation).

Participants ($N$ = 130, 40.0% female, $M_{age}$ = 39.39, $SD_{age}$ = 12.66) were recruited on MTurk and, as in Study 1, introduced to a political candidate named Matthew Johnson. In this study, Mr. Johnson was *always* described as taking an absolute stance on honesty during a television interview, but we varied the broadness of the absolute honesty stance such that Mr. Johnson made a broad claim against lying (i.e., *I believe that it is never okay to lie*) or a specific claim against lying about accepting political action committee (PAC) donations (i.e., *I believe that it is never okay to lie about accepting political action committee (PAC) donations*). After the television interview, Matthew Johnson was described as lying about accepting PAC donations during his campaign, using the same stimuli as in Studies 1-2 and Supplemental Study 4. Therefore, Mr. Johnson either violated the exact same type of honesty stance he had just preached about on television (never lying about accepting PAC donations) or violated a broader, more generalized honesty stance (never lying).

After learning this information, participants indicated their intentions of voting for Mr. Johnson and evaluated Mr. Johnson on perceptions of morality and hypocrisy. We were interested to see whether more specific stances were perceived as more hypocritical, and as a result, were less likely to reflect good moral character.

***Dependent Variables***

Participants indicated their likelihood of voting for Mr. Johnson, as well as the morality ($\alpha$ = .95) and hypocrisy ($\alpha$ = .88) of Mr. Johnson using the same scales as in Studies 1 and 2.

**Results**

A series of independent t-tests were conducted to examine the effects of broadness (broad absolute proclamation vs. specific absolute proclamation) on voting intentions and attitudinal evaluations.

***Voting Intentions***

There was no significant effect of broadness on voting intentions ($t(128) = 0.88$, $p = .381$, $d = 0.15$). Participants were not more or less likely to vote for the candidate who took the broad absolute honesty stance ($M = 2.78$, $SD = 1.19$) than the candidate who took the specific absolute honesty stance ($M = 2.59$, $SD = 1.28$).

***Morality***

There was no significant effect of broadness on moral evaluations ($t(128) = -0.47$, $p = .642$, $d = -0.08$). Participants viewed the candidate who took the broad absolute honesty stance as similarly moral ($M = 2.40$, $SD = 1.02$) to the candidate who took the specific absolute honesty stance ($M = 2.48$, $SD = 1.15$).

***Hypocrisy***

There was no significant main effect of broadness on hypocrisy evaluations ($t(128) = -0.23$, $p = .820$, $d = -0.04$). The candidate who took the broad absolute honesty stance ($M = 5.76$, $SD = 0.84$) and the candidate who took the specific absolute honesty stance ($M = 5.80$, $SD = 1.02$) were seen as similarly hypocritical.

## 2.7 Supplemental Study 7

### Abbreviated Methods

In Supplemental Study 7, we conducted an almost identical study as Study 2 reported in the main manuscript. Participants learned about a hypothetical candidate, Matthew Johnson, who took either an absolute honesty stance, flexible honesty stance, or no stance and then lied. Unlike Study 2, participants only evaluated the candidate at one time. We extended Study 2 by including additional measures of potential mechanisms in this study. We examine the perceived social

benefit of moral proclamations, as in Study 5, and assess perceptions of lying behavior and honesty ideals, as in Supplemental Study 3 and 4.

Participants ($N$ = 265, 40.8% female, $M_{age}$ = 39.11, $SD_{age}$ = 12.04) were assigned to a proclamation condition (absolute vs. flexible vs. control) in a between-subjects design. All participants read a television transcript of an interview between a hypothetical political candidate, Matthew Johnson, and a television reporter, in which they learned about the candidate's honesty stance. The stimuli were identical to Study 2 in the main manuscript.

After reading the interview, participants continued to learn about the candidate and learned that Matthew Johnson lied about his campaign financing. Participants learned all this information before completing any evaluations, as in Study 1. After learning about Matthew Johnson, participants indicated their likelihood of voting for this candidate and evaluated Matthew Johnson on morality, hypocrisy, and future honesty. We were interested to see whether the flexible honesty stance was costly relative to both the absolute honesty stance and to taking no stance at all.

We also explore additional mechanisms in this study. The future honesty mechanism is intended to capture participant perceptions of Matthew Johnson's valuation of honesty and likelihood of engaging in honest behavior. However, it is also possible that participants think that a candidate can value honesty and strive to be honest, but also lie regularly if the stance of *never* lying is an impossible standard. Therefore, we included exploratory measures meant to capture participant perceptions of Matthew Johnson's *actual* levels of lying and beliefs about the acceptability of lying in everyday life (as in Studies S3-S4), which could be distinct from participant perceptions of Matthew Johnson's future honesty. We also measure the perceived social benefit of the proclamations for participants in the absolute and flexible honesty

conditions, as in Study 5. Descriptive statistics for these exploratory attitudinal measures are reported in Table SI-22.

***Dependent Variables***

Participants indicated how likely they would be to vote for the candidate, and evaluated the candidate's morality, hypocrisy, and future honesty using the same measures as in Study 2 ($\alpha s \geq .74$). The social benefit of the proclamation was measured for participants in the absolute and flexible honesty condition using the same two items from Study 5, which we combined due to high agreement ($r(160) = .87$, $p < .001$).

**Lying Behavior.** We include the same measure as in Supplemental Studies 3-4, with a slightly modified scale, examining how often participants thought Matthew Johnson ***actually*** lies. In this measure, higher scores reflect more frequent assumptions of lying ("How often do you think Matthew Johnson **actually lies**?" 1 = never ever to 7 = always).

**Honesty Ideals.** We include the same measure of honesty ideals as in Supplemental Studies 3-4, with a slightly modified scale, which reflects participants' perceptions of how often Matthew Johnson thought it was ***actually okay*** to lie. In this measure, higher scores reflect less strict honesty ideals ("How often do you think Matthew Johnson **actually thinks it is okay to lie**?" 1 = never ever to 7 = always).

**Results**

We report the results of a one-way ANOVA, using proclamation as a factor, on each of our dependent variables. We followed up significant main effects of proclamation with independent-samples t-tests to compare differences in evaluations between conditions.

***Voting Intentions***

There was a significant main effect of proclamation ($F(2, 262) = 8.81$, $p < .001$, $\eta_p^2 = .06$) on voting intentions. Participants were less likely to vote for the candidate who endorsed absolute honesty than the control candidate ($t(184) = -2.54$, $p = .012$, $d = -0.37$), consistent with past work on the social costs of hypocrisy. Participants were also directionally less likely to vote for the candidate who endorsed flexible honesty than the candidate who endorsed absolute honesty, though this effect was not significant at the $p < .05$ level ($t(160) = 1.60$, $p = .111$, $d = 0.25$).

***Morality***

There was a main effect of proclamation on morality ratings ($F(2, 262) = 3.07$, $p = .048$, $\eta_p^2 = .02$). Although the control candidate was (directionally) considered to be the most moral, this candidate was not seen as significantly more moral than the candidate who endorsed absolute honesty ($t(184) = -0.56$, $p = .576$, $d = -0.08$). The candidate who endorsed flexible honesty was seen as directionally, but only marginally significantly less moral than the candidate who endorsed absolute honesty ($t(160) = 1.79$, $p = .075$, $d = 0.28$).

***Hypocrisy***

There was a main effect of proclamation ($F(2, 262) = 13.31$, $p < .001$, $\eta_p^2 = .09$), such that participants viewed the candidate who endorsed absolute honesty as more hypocritical than the control candidate ($t(184) = 3.74$, $p < .001$, $d = 0.55$) and the candidate who endorsed flexible honesty ($t(160) = 5.17$, $p < .001$, $d = 0.81$). The control candidate was also seen as directionally, but only marginally significantly, more hypocritical than the candidate who endorsed flexible honesty ($t(180) = 1.71$, $p = .089$, $d = 0.25$).

***Future Honesty***

There was also a main effect of proclamation ($F(2, 262) = 9.91$, $p < .001$, $\eta_p^2 = .07$) on future honesty. The candidate who endorsed flexible honesty was viewed as less likely to engage in honesty than the candidate who endorsed absolute honesty ($t(160) = 4.29$, $p < .001$, $d = 0.67$). and the control candidate ($t(180) = 3.67$, $p < .001$, $d = 0.54$). The candidate who endorsed absolute honesty and the control candidate were not viewed significantly differently ($t(184) = 0.48$, $p = .633$, $d = 0.07$). Thus, candidates endorsing absolute honesty were not seen as any less likely to engage in future honesty than the default assumption (i.e., the control candidate), despite their hypocrisy.

***Social Benefit of the Proclamation***

There was a main effect of proclamation on perceived social benefit of the proclamation ($F(1, 160) = 221.44$, $p < .001$, $\eta_p^2 = .58$). The absolute honesty proclamation ($M = 4.61$, $SD = 1.37$) was viewed as having a greater societal benefit than the flexible honesty proclamation ($M = 1.84$, $SD = 0.95$). Participants in the control condition did not complete this measure, since control candidates did not make any proclamations about honesty.

***Lying Behavior***

There was a main effect of proclamation ($F(2, 262) = 5.68$, $p = .004$, $\eta_p^2 = .04$). Participants thought that the candidate who made the flexible proclamation ($M = 4.97$, $SD = 0.96$) lied more frequently than the candidate who made the absolute proclamation ($M = 4.51$, $SD = 0.99$) or the candidate in the control condition ($M = 4.50$, $SD = 1.15$).

***Honesty Ideals***

There was a main effect of proclamation, ($F(2, 262) = 8.61$, $p < .001$, $\eta_p^2 = .06$). The candidate who made the flexible proclamation was seen as having more modest honesty ideals ($M = 5.15$, $SD = 1.00$) relative to the candidate who made the absolute proclamation ($M = 4.47$,

*SD* = 1.30) and the control candidate (*M* = 4.49, *SD* = 1.27). Making the absolute proclamation, though perceived as more hypocritical, still sent a signal of more ambitious honesty ideals than the flexible proclamation.

**Table SI-22**

*Descriptive Statistics in Supplemental Study 7*

| **Variable** | **Proclamation** | | |
|---|---|---|---|
| | Absolute | Flexible | Control |
| Voting intentions | 2.81 (1.24) | 2.48 (1.35) | 3.33 (1.50) |
| Morality | 2.70 (1.07) | 2.39 (1.10) | 2.79 (1.12) |
| Hypocrisy | 5.70 (0.93) | 4.78 (1.31) | 5.10 (1.20) |
| Future honesty | 3.05 (0.89) | 2.45 (0.89) | 2.98 (1.02) |
| Social benefit of proclamation | 4.61 (1.37) | 1.84 (0.95) | - |
| Lying behavior | 4.51 (0.99) | 4.97 (0.96) | 4.50 (1.15) |
| Honesty ideals | 4.47 (1.30) | 5.15 (1.00) | 4.49 (1.27) |

***Multiple Mediation Additional Results***

We conducted a multiple mediation analysis using a bootstrapping procedure with the lavaan package in R (Rosseel, 2012) with proclamation entered as the independent variable (1 = flexible, 0 = absolute), hypocrisy, future honesty, and social benefit of the proclamation entered as simultaneous mediators, and moral evaluations entered as the dependent variable. We did not include participants from the control condition in this analysis, consistent with Study 2 in the main text, and because our social benefit mechanism was not measured in the control condition.

Results indicated that the flexible honesty proclamation signaled lower hypocrisy and less future honesty than the absolute honesty proclamation, and both of these inferences predicted moral evaluations. Decreased perceptions of future honest behavior were correlated with more negative moral evaluations, whereas decreased perceptions of hypocrisy were correlated with more positive moral evaluations. However, we did not find evidence that the social benefit of the proclamation mediated the relationship between proclamation and moral

evaluations in this analysis (Indirect Effect through Hypocrisy = 0.253, 95% CI = [0.13, 0.40]; Indirect Effect through Future Honesty = -0.383, 95% CI = [-0.59, -0.19]; Indirect Effect through Social Benefit of the Proclamation = -0.254, 95% CI = [-0.53, 0.01]). As in Study 1, the positive indirect effect through future honesty was larger than negative indirect effect through hypocrisy, which helps to explain why flexible, but realistic, stances lead to more negative moral judgments than absolute, yet hypocritical, stances.

We also conducted a second multiple mediation analysis with proclamation entered as the independent variable (0 = absolute, 1 = flexible) and voting intentions as the dependent variable. Hypocrisy, future honesty, and the perceived social benefit of the proclamation were entered as simultaneous mediators. We again exclude participants in the control condition in this analysis, since they never received the question about the social benefit of the proclamation.

Results indicate a significant indirect influence through hypocrisy (Indirect Effect through Hypocrisy = 0.216, 95% CI = [0.05, 0.40]), future honesty (Indirect Effect through Future Honesty = -0.325, 95% CI = [-0.55, -0.13]), and the perceived social benefit of the proclamation (Indirect Effect through Social Benefit of the Proclamation = -0.653, 95% CI = [-1.10, -0.25]). Similar to the results in the model on moral evaluations, the flexible proclamation lowered both perceptions of hypocrisy and future honesty, which had opposite effects on voting intentions. Unlike the model on moral evaluations, we find support in this analysis that the perceived social benefit of the proclamation also mediates the relationship between proclamation and voting intentions. The social benefit of the flexible proclamation was seen as less valuable than the social benefit of the absolute proclamation, which positively correlated with intentions of voting for Matthew Johnson. Combined with the mediation results throughout our studies, we find inconsistent support in favor of the social benefit of the proclamation as a mechanism, but

find consistent support for perceptions of future honest behavior as a mechanism that explains preferences for communicators endorsing absolute honesty.

## 3. Pilot Study 4

The goal of Pilot Study 4 was to assess reactions towards communicators who endorsed absolute and flexible honesty stances with caveats. We thought it was possible that adding caveats to the flexible stance (i.e., "*It is sometimes okay to lie if you have a good reason*") might make communicators endorsing the flexible stance seem more moral or likely to engage in future honesty, or that adding caveats to the absolute stance might convey flexibility in a more palpable manner (i.e., "*It is never okay to lie, except if you have a good reason*").

Since the majority of participants in our studies did not endorse absolute honesty in private (see Pilot Studies 1-3 and Studies 4-5 in the main manuscript), and were relatively tolerant of hypocrisy, we thought it might be the case that people think it is never okay to lie most of the time, but it is acceptable to lie for certain reasons (i.e., prevent harm). Therefore, communicators might seem more moral and likely to be honest if they explicitly mention these exceptional circumstances. Pilot Study 4 was an exploratory investigation of this question.

**Methods**

Participants ($N$ = 296, 40.9% female, $M_{age}$ = 39.02, $SD_{age}$ = 10.57) were recruited on MTurk for this pilot. We introduced participants to another past study participant who had taken a survey on beliefs and preferences. Our participants saw screenshots of the past participant's survey responses, which manipulated the honesty stance, using a similar paradigm as in Studies 3-5 in the main manuscript. Participants were randomly assigned to one of 10 conditions. The

honesty stance varied across the conditions, but all other presented preferences (responses in the survey) were held constant. The stances were as follows:

**[Original stances]**
**Absolute:** *Its never okay to lie*
**Ambitious**: *Its rarely okay to lie*
**Flexible:** *Its sometimes okay to lie*
**[Absolute stances + caveats]**
**Absolute + exceptional caveat**: *Its never okay to lie, except in really exceptional circumstances*
**Absolute + good reason caveat**: *Its never okay to lie, except if you have a good reason*
**Absolute + harm caveat:** *Its never okay to lie, except if it prevents harm*
**[Flexible + caveat]**
**Flexible + exceptional caveat**: *Its sometimes okay to lie, but only in really exceptional circumstances*
**Flexible + good reason caveat**: *Its sometimes okay to lie if you have a good reason*
**Flexible + harm caveat**: *Its sometimes okay to lie if it prevents harm*
**[Control]**
**Control**: -- [no honesty stance]

After learning this information, participants evaluated the communicator in terms of morality and future honesty, using the same scales as in Studies 3-5 ($\alpha s \geq .79$).

We also included a measure at the end of the survey asking participants in conditions 1-9 (excluding participants in the control condition) to indicate the extent to which they agreed with the communicator's honesty stance ("To what extent do you agree with [randomly assigned communicator]'s attitude on lying?"), from 1 = Strongly disagree to 7 = Strongly agree.

**Results**

***Generalized Proclamations***

First, in an ANOVA comparing evaluations of communicators who made the generalized absolute vs. ambitious vs. flexible stance on honesty, consistent with those used in Studies 4-5 in the manuscript, we replicate our previous results by finding more positive moral evaluations and increased expectations of future honesty for communicators who take absolute honesty statements than those who take flexible honesty statements (absolute vs. flexible, morality: $t(58)$

= 5.00, $p < .001$, $d = 1.29$; absolute vs. flexible, future honesty: $t(58) = 7.36$, $p < .001$, $d = 1.90$). In this pilot, the absolute stance was also viewed more positively than the ambitious stance (absolute vs. ambitious, morality: $t(56) = 2.25$, $p = .029$, $d = 0.59$; absolute vs. ambitious, future honesty: $t(56) = 2.42$, $p = .019$, $d = 0.63$). The control communicator was viewed as less likely to engage in future honesty (control vs. absolute, future honesty: $t(56) = 2.66$, $p = .010$, $d = 0.70$) and similarly moral (absolute vs. control, morality: $t(56) = 1.72$, $p = .090$, $d = 0.45$) as the communicator endorsing absolute honesty, and as *more* moral and likely to engage in future honesty than the communicator endorsing the flexible stance (flexible vs. control, morality: $t(58) = -3.80$, $p < .001$, $d = -0.98$; flexible vs. control, future honesty: $t(58) = -5.50$, $p < .001$, $d = -1.42$), consistent with our theoretical model.

***Honesty Caveats***

When examining caveats, as mentioned in the main manuscript, we found that adding a caveat explaining when and why one might lie (for positive reasons) did make communicators taking the flexible honesty stance seem more moral and more likely to engage in future honesty at times. Specifically, the caveats of preventing harm ($t(59) = -2.54$, $p = .014$, $d = -0.65$) and exceptional circumstances ($t(58) = -2.18$, $p = .033$, $d = -0.56$) made the communicator seem significantly more moral than when endorsing general flexibility alone (these means are bolded in Table SI-23). Adding the caveat of having a "good reason" did not change moral evaluations of communicators relative to the generalized flexible stance ($t(60) = -0.44$, $p = .661$, $d = -0.11$). Furthermore, only the caveat of "exceptional circumstances" made communicators seem more likely to engage in future honesty relative to the generalized flexible stance ($t(58) = -2.77$, $p = .007$, $d = -0.72$).

Adding caveats to the absolute stance, on the other hand, did not make boost moral or future honesty evaluations. In fact, communicators who endorsed the generalized absolute stance were seen as significantly more moral ($t(57) = 3.01$, $p = .004$, $d = 0.78$) and more likely to engage in future honesty ($t(57) = 3.40$, $p = .001$, $d = 0.88$) than those who endorsed the absolute stance with the good reason caveat. Communicators who endorsed the generalized absolute stance were not seen as any different in morality than those who took this stance with the caveat of preventing harm ($t(56) = 1.31$, $p = .196$, $d = 0.34$) or having exceptional circumstances for lying ($t(56) = 1.97$, $p = .054$, $d = 0.52$). Yet, even in these cases, the generalized absolute honesty stance still led to greater expectations of future honesty relative to the absolute stance with both caveats of preventing harm ($t(56) = 2.25$, $p = .029$, $d = 0.59$) or having exceptional circumstances for lying ($t(56) = 2.40$, $p = .020$, $d = 0.63$; significant differences are bolded in Table SI-23).

Although caveats did make the flexible stance seem better at times, the generalized absolute honesty stance still led to greater moral evaluations than the flexible stances even with these specific caveats (absolute vs. flexible + exceptional, $t(56) = 2.89$, $p = .006$, $d = 0.76$; absolute vs. flexible + harm: $t(57) = 2.49$, $p = .016$, $d = 0.65$).

***Generalized Proclamations***

We were interested to see if participants strongly agreed with the generalized absolute honesty stance in this pilot study. We compared participant agreement with the generalized absolute honesty stance to agreement with the flexible honesty stance with the caveats of preventing harm and needing exceptional circumstance to lie, since these were the caveats that elevated impressions of communicators relative to the generalized flexible stance. We find no significant differences of agreement in these cases (absolute vs. flexible + preventing harm: $t(57) = -0.65$, $p = .517$, $d = -0.17$; absolute vs. flexible + exceptional circumstances: $t(56) = -0.71$, $p =$

.478, $d$ = -0.19). Participants had similar agreement for generalized absolute honesty as flexible honesty with these two caveats, despite evaluating the communicator endorsing absolute honesty more positively.

**Table SI-23**

*Comparing Honesty Stances with Caveats*

| **Stance** | **Morality** | | **Future Honesty** | |
|---|---|---|---|---|
| | ***M*** | ***SD*** | ***M*** | ***SD*** |
| Absolute | 5.97 | 1.02 | 5.86 | 1.02 |
| Absolute + exceptional | 5.39 | 1.19 | **5.19** | **1.11** |
| Absolute + good reason | **5.10** | **1.17** | **4.94** | **1.06** |
| Absolute + harm | 5.59 | 1.18 | **5.25** | **1.05** |
| Flexible | 4.19 | 1.63 | 3.77 | 1.17 |
| Flexible + exceptional | **5.05** | **1.38** | **4.63** | **1.22** |
| Flexible + good reason | 4.34 | 0.98 | 3.87 | 0.84 |
| Flexible + harm | **5.18** | **1.38** | 4.23 | 0.85 |
| Ambitious | 5.43 | 0.79 | 5.27 | 0.85 |
| Control | 5.52 | 0.95 | 5.22 | 0.82 |

*Notes.* Bold indicates a significant difference of adding the caveat relative to the generalized stance (i.e., Flexible vs. Flexible + exceptional: the exceptional caveat leads to significantly more positive evaluations in both morality and future honesty). When significant, adding caveats to the flexible stance directionally increased morality and future honesty evaluations relative to the flexible generalized stance, but adding caveats to the absolute stance *decreased* morality and future honesty evaluations relative to the absolute generalized stance (i.e., Absolute vs. Absolute + exceptional: the exceptional caveat leads to significantly *less* perceptions of future honesty).

## 4. Links to Pre-registrations for Studies 1-5 and Supplemental Studies 1-4

Study 1 – Moral Stances in the Political Context
http://aspredicted.org/blind.php?x=w2fj5s

Study 2 – Moral Stances in the Political Context Over Time
https://aspredicted.org/KJ4_MS3

Study 3 –Moral Stances in the Laboratory Over Time
https://aspredicted.org/COH_AFM

Study 4 – Prosocial Lies and Selfish Truths
https://aspredicted.org/ZBJ_SIL

Study 5 – Selfish Lies and Prosocial Truths
https://aspredicted.org/GXM_XSQ

Supplemental Study 1:
https://aspredicted.org/blind.php?x=uf42y8

Supplemental Study 2:
http://aspredicted.org/blind.php?x=87ef8k

Supplemental Study 3:
http://aspredicted.org/blind.php?x=mk2fa4

Supplemental Study 4:
https://aspredicted.org/blind.php?x=6s7sx4

Supplemental Study 7:
https://aspredicted.org/ULB_NLQ