Research Article

Psychological Science
1–13

DOI: 10.1177/0956797616685771
journals.sagepub.com/home/pss

Why Do We Hate Hypocrites? Evidence for a Theory of False Signaling

Jillian J. Jordan[1], Roseanna Sommers[1], Paul Bloom[1], and David G. Rand[1,2,3]

[1]Department of Psychology, [2]Department of Economics, and [3]School of Management, Yale University

Abstract

Why do people judge hypocrites, who condemn immoral behaviors that they in fact engage in, so negatively? We propose that hypocrites are disliked because their condemnation sends a false signal about their personal conduct, deceptively suggesting that they behave morally. We show that verbal condemnation signals moral goodness (Study 1) and does so even more convincingly than directly stating that one behaves morally (Study 2). We then demonstrate that people judge hypocrites negatively—even more negatively than people who directly make false statements about their morality (Study 3). Finally, we show that "honest" hypocrites—who avoid false signaling by admitting to committing the condemned transgression—are not perceived negatively even though their actions contradict their stated values (Study 4). Critically, the same is not true of hypocrites who engage in false signaling but admit to unrelated transgressions (Study 5). Together, our results support a false-signaling theory of hypocrisy.

Keywords

moral psychology, condemnation, vignettes, deception, social signaling, open data, open materials

Received 5/30/16; Revision accepted 12/2/16

Consider the hypocrite—someone who condemns the moral failings of other people but behaves badly him- or herself. Many commentators have remarked on the "peculiarly repulsive" nature of hypocrisy (Shklar, 1984, p. 57). The degree to which hypocrites are disliked cannot be explained by their transgressions alone: What makes hypocrites especially bad is that they both commit a transgression and condemn it. But why is this combination so objectionable? After all, speaking out against immorality is normally seen as laudable. It enforces norms and encourages moral behavior (Berkowitz & Walker, 1967; Feinberg, Willer, & Schultz, 2014; Feinberg, Willer, Stellar, & Keltner, 2012), such that failing to condemn transgressions has been characterized as second-order free-riding (Yamagishi, 1986). Arguably, then, people should not be so resentful toward hypocrites: They may fail to achieve their moral aspirations, but at least they oppose bad behavior.

Previous research has investigated the psychology of hypocrites, including how hypocrites justify their behavior (Batson, Thompson, Seuferling, Whitney, & Strongman, 1999) and address aversive cognitive dissonance (Aronson, Fried, & Stone, 1991). Relatively little work has examined how hypocrisy is perceived by other people (but see Alicke, Gordon, & Rose, 2013; Barden, Rucker, & Petty, 2005).

One reason hypocrisy is perceived negatively may be that it involves inconsistency between words and deeds, which people tend to dislike (Tedeschi, Schlenker, & Bonoma, 1971). Another possibility is that hypocrites may be seen as unable to resist the temptation to transgress—another negative quality (Righetti & Finkenauer, 2011). Furthermore, hypocrites may be seen as more intentionally immoral than people who behave badly without condemning such behavior (Cushman, Young, & Hauser, 2006): Their condemnation demonstrates that they understand the wrongfulness of their actions.

Here, we propose a different hypothesis, based on the idea of false signaling. We suggest that hypocrites are

Corresponding Author:
Jillian J. Jordan, Yale University–Psychology, 1 Prospect St., New Haven, CT 06511
E-mail: jillian.jordan@yale.edu

disliked because they use their condemnation to mislead other people about their moral behavior.

As a matter of logic, there is nothing dishonest about both taking an action and condemning it. But engaging in moral condemnation may be perceived as communicating information about one's future behavior (Baumeister, Zhang, & Vohs, 2004). The idea that condemnation may signal moral goodness is consistent with evidence that people who punish selfish players in economic games are seen as more trustworthy than people who choose not to punish (Barclay, 2006; Horita, 2010; Jordan, Hoffman, Bloom, & Rand, 2016; Nelissen, 2008; Raihani & Bshary, 2015a, 2015b). We thus hypothesize that hypocrites inspire moral outrage because they dishonestly signal their moral goodness—that is, their condemnation of immoral behavior signals that they are morally upright, but they fail to act in accordance with these signals.

This theory of false signaling helps explain why hypocrites are often regarded as liars (McKinnon, 1991). But it also predicts that hypocrites may be seen as worse than people who falsely claim to behave morally, whom we refer to as *direct liars*: In cases in which moral condemnation acts as a more persuasive signal than directly claiming to behave morally would, hypocrites are actually more misleading than direct liars. Furthermore, a hypocrite's false signals may be more destructive than a liar's false statements (e.g., by earning the hypocrite undeserved trust or by manipulating other people into following the hypocrite's professed standards) and may come at the expense of other people (e.g., because condemnation tarnishes the reputation of the condemned; Williams, Forgas, & von Hippel, 2005). Liars, by contrast, avoid moral condemnation and are thus less likely to malign or shame other people.

Finally, a key prediction of our false-signaling theory is that *honest hypocrites*, who admit to committing the acts they condemn, should not be judged negatively for behaving hypocritically because their condemnation has been stripped of any signaling function. In other words, if hypocrites are disliked because of their false signaling, a hypocrite who admits to transgressing should be forgiven—insofar as this admission cancels any false signals. In the five studies reported here, we tested these predictions.

Study 1

We began with the hypothesis that moral condemnation is treated as a signal that one will behave morally in the future. According to this theory, individuals who condemn others' transgressions should be perceived as less likely to commit those transgressions, and as overall more moral, than individuals who have not conversed about the transgressions. But condemnation should have this signaling effect only in the absence of more direct information about the condemner's morality. If condemners are perceived positively because their condemnation signals that they will behave morally, their condemnation should no longer matter when a more informative indicator of moral behavior is available.

Method

Design. To test these predictions, we presented subjects with vignettes and asked them to evaluate target characters in the vignettes. In a 2 × 2 between-subjects design, we manipulated whether the targets engaged in moral condemnation or not and whether subjects had direct information about the targets' moral behavior or not. We predicted that subjects would evaluate targets who engaged in moral condemnation more positively than those who did not, but only in the absence of direct information about the targets' moral behavior.

Subjects. We recruited subjects online using Amazon Mechanical Turk (MTurk). We predicted an interaction between our two independent variables but did not have a clear prediction for what the effect size for this interaction would be. Thus, we precommitted to recruiting 800 subjects (n = 200 per condition), which is our standard protocol for between-subjects designs on MTurk when an interaction is predicted. A total of 798 people actually completed the survey, which required them to evaluate all the target characters and answer all the comprehension questions correctly (see the Procedure section for details). However, we could not analyze the responses of the first 135 subjects who completed the survey because of a technical error in how the survey was programmed (we corrected this error before the remaining subjects participated). We analyzed the data of all remaining subjects who had unique IP (Internet protocol) addresses (to avoid duplicate respondents). Our final sample consisted of 619 subjects (mean age = 31 years, 59% male).

Procedure. In each of our vignettes, we asked subjects to imagine that they belonged to a social group in which a particular moral transgression was possible (e.g., a track team whose members might use forbidden performance-enhancing drugs). Subjects were then told about two members of the social group: the *target* (whom subjects would later evaluate) and the *other person* (whom subjects would not evaluate), both of whom were described neutrally (not using these terms).

We then manipulated whether subjects received direct positive information about the moral behavior of the target. In the *no-information* condition, subjects were given no information about the moral behavior of the target or the other person. In the *good-information* condition, subjects received direct, positive information about the

moral behavior of the target (but not the other person): They were told that the target recently behaved morally (e.g., did not use drugs in his or her last athletic competition).

In addition, we manipulated whether the target engaged in moral condemnation of a wrongdoer. In the *target-condemns* condition, we asked subjects to imagine having a dialogue with the target in which the target mentioned that a mutual acquaintance recently behaved immorally (e.g., used drugs at an athletic competition) and expressed strong disapproval of this acquaintance's behavior. In the *other-condemns* condition, subjects were told to imagine having the same dialogue, but with the other person instead of with the target. Thus, in all conditions, subjects read the same description of condemnation, but in the *target-condemns* condition, the target engaged in the condemnation, whereas in the *other-condemns* condition, the other person did (and the target was absent from the conversation).

For example, following is the full text for one scenario about performance-enhancing drugs. In this scenario, Brian is the target character, and Sam is the other person.

> Imagine that you are an athlete on a track team. Recently, your coach has become concerned that members of the team are using an illegal performance-enhancing drug called Vitronil. Vitronil use threatens your team's eligibility to compete, and gives individual athletes unfair advantages.

In the no-information condition, the scenario continued as follows:

> Two of your teammates are named **Brian** and **Sam**. You know nothing about if **Brian** uses Vitronil. You also know nothing about if **Sam** uses Vitronil.

In the good-information condition, the scenario instead continued with

> Two of your teammates are named **Brian** and **Sam**. You overheard another member of the track team saying that **Brian** did not use Vitronil at his last track competition. In contrast, you know nothing about if **Sam** uses Vitronil.

The scenario concluded as follows in the target-condemns condition:

> One day, you are having a conversation with **Brian**. You tell them a story about a mutual acquaintance, **Mark**, who is a competitive swimmer. After you finish your story, **Brian** mentions that he heard that **Mark** got caught using Vitronil right before an important swim meet. In telling his story, **Brian** expresses strong disapproval of Vitronil use.

The closing passage in the other-condemns condition was the same except that all references to Brian were changed to references to Sam:

> One day, you are having a conversation with **Sam**. You tell them a story about a mutual acquaintance, **Mark**, who is a competitive swimmer. After you finish your story, **Sam** mentions that he heard that **Mark** got caught using Vitronil right before an important swim meet. In telling his story, **Sam** expresses strong disapproval of Vitronil use.

After reading each vignette, subjects evaluated the target, using Likert scales from 1 to 7. To measure expectations of the targets' future moral behavior, we used four items that ranged in their specificity. Subjects rated the targets on their likelihood of committing the relevant moral transgression (e.g., for the scenario just quoted, "How likely do you think Brian is to use Vitronil in the future?"), their trustworthiness in the specific domain relevant to the vignette (e.g., "How much would you trust Brian as a competitor on your team?"), their general trustworthiness (e.g., "How much would you generally trust Brian across contexts?"), and their likeability (e.g., "How much do you like Brian?"). For the question about the likelihood of transgressing, the scale ranged from *very unlikely* to *very likely*, and for the questions about trustworthiness and likeability, it ranged from *very little* to *very much*.

Each subject was presented with four vignettes, describing (a) a track team whose members could use performance-enhancing drugs (as in the vignette just quoted), (b) a chemistry course in which students could cheat on take-home exams, (c) a work organization in which employees could fail to meet deadlines on team projects, and (d) a hiking club whose members frequently dated each other and could engage in infidelity. The four vignettes were presented in random order. To reduce noise, we matched all characters in the athletic, academic, and work scenarios to the subject's gender and made all characters in the romantic scenario of the opposite gender. The full text of all the vignettes is in the Supplemental Material available online.

Immediately after reading each vignette, subjects answered four comprehension questions. If subjects answered a question incorrectly, they were not allowed to continue participating (i.e., the evaluation questions were presented, but the survey automatically ended before the next vignette was presented). A total of 85.9% of subjects showed perfect comprehension on all the vignettes, and the percentage of subjects with perfect comprehension did not differ significantly across conditions, $\chi^2(3, N = 747) = 2.26$, $p = .521$.

We found high interitem reliability among the four individual dependent measures of evaluation of the

targets ($\alpha = .90$); thus, we averaged these ratings to create one composite scale representing positive evaluations. We report analyses of this composite measure, but we also investigated possible differences among the individual dependent measures and found that the results were largely robust across these measures (see the Supplemental Material). To compute the composite ratings, we first reverse-coded the rating of the target's likelihood of transgressing and then took the mean rating across the four variables.

Results

To test our predictions, we conducted a 2 (condemnation condition: target condemns vs. other condemns) × 2 (information condition: good information vs. no information) analysis of variance (ANOVA) predicting mean positive evaluations of the targets across the vignettes (see Fig. 1). We found a significant main effect of information condition, $F(1, 615) = 137.93$, $p < .001$, $\eta_p^2 = .183$; subjects evaluated targets more positively in the good-information condition ($M = 5.28$, $SD = 0.83$) than in the no-information condition ($M = 4.52$, $SD = 0.80$). This result served as a manipulation check, demonstrating that direct positive information about the target's moral behavior was perceived as a clear indication of moral goodness.

We also found a significant main effect of condemnation condition, $F(1, 615) = 13.20$, $p < .001$, $\eta_p^2 = .021$; subjects evaluated targets more positively when the target engaged in condemnation ($M = 5.01$, $SD = 0.91$) than when the other person engaged in condemnation ($M = 4.81$, $SD = 0.87$). This result confirmed our hypothesis that moral condemnation serves as a signal of moral goodness.

Finally, we found a significant interaction, $F(1, 615) = 8.51$, $p = .004$, $\eta_p^2 = .014$; the target's use of condemnation had a larger effect in the no-information condition than in the good-information condition. Specifically, when subjects were given no information about the target's behavior, they evaluated the target significantly more positively when he or she condemned the transgression ($M = 4.73$, $SD = 0.94$) than when the other party condemned the transgression ($M = 4.31$, $SD = 0.55$), mean difference = 0.42, 95% confidence interval (CI) = [0.25, 0.59], $t(301) = 4.77$, $p < .001$, $d = 0.55$. However, in the good-information condition, we found no significant difference between the target-condemns condition ($M = 5.30$, $SD = 0.79$) and the other-condemns condition ($M = 5.25$, $SD = 0.87$), mean difference = 0.05, 95% CI = [−0.14, 0.23], $t(314) = 0.49$, $p = .622$, $d = 0.06$. This result confirmed our hypothesis that observers rely on a person's statements of condemnation as a signal of moral goodness only when they lack direct information about the person's moral behavior.

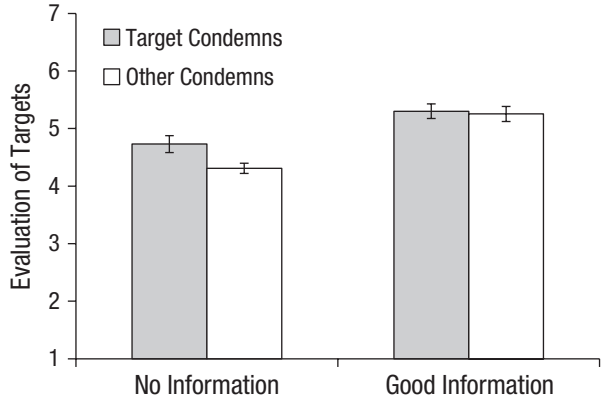


Fig. 1. Results from Study 1: mean composite evaluation of the targets as a function of condemnation condition and information condition. Error bars represent 95% confidence intervals.

We note that the null effect of the target expressing condemnation in the good-information condition does not appear to be a ceiling effect. In the good-information condition, the mean composite evaluation (5.28) was substantially below the scale's ceiling (7), and subjects rarely used the ceiling value (only 15.5% of responses to the evaluative questions were a "7").

Thus, our data support our prediction that a person's condemnation of a transgression serves as a signal of moral behavior—when direct information about the condemner's behavior is unavailable. This suggests that condemnation is viewed positively because it signals moral behavior.

Study 2

Study 1 sheds light on why hypocrites are typically thought of as liars: If condemnation signals morality, then hypocrites mislead other people. How, then, do hypocrites, who use condemnation to imply (falsely) that they behave morally, compare with outright liars, who directly state (falsely) that they behave morally? Our theory predicts that hypocrites might be seen as worse than liars in situations in which condemnation is perceived as a stronger signal than a direct statement—and thus their deception is more misleading. Thus, in Study 2 we compared the signaling strength of condemnation of transgressions and direct statements of moral behavior.

Method

Design. The design of Study 2 was similar to that of Study 1, with just a few modifications. We again asked subjects to evaluate target characters in vignettes. In a 2 × 2 between-subjects design, we manipulated whether the target sent a signal of moral goodness or not and whether

that signal was moral condemnation (of another person's transgression) or a direct statement (that the target did not engage in that transgression). In all conditions, we provided subjects with no direct information about the target's moral behavior (as in the no-information condition of Study 1), because this is the condition in which we found condemnation to have a significant effect on evaluations.

Subjects. We again recruited subjects using MTurk. As in Study 1, we precommitted to recruiting 800 subjects (n = 200 per condition); a total of 838 people actually completed the survey. We again analyzed the data of all subjects who had unique IP addresses, evaluated all the targets, and answered all the comprehension questions correctly. Our final sample consisted of 803 subjects (mean age = 31 years, 59% male).

Procedure. The vignettes described the same social groups as in Study 1 and again introduced the target and the other person. In each vignette, subjects were told to imagine having a conversation about which members within the social group typically engaged in the transgression in question (e.g., used drugs in athletic competitions). In the *target-signals* condition, subjects were told to imagine having this conversation with the target, whereas in the *other-signals* condition, subjects were told to imagine having it with the other person. Within this conversation, in the *condemnation* condition, subjects were told that the person they were talking to (the target or other person, depending on condition) emphasized his or her disapproval of the transgression. In the *direct-statement* condition, subjects were instead told that the person they were talking to emphasized that he or she did not engage in the transgression. Then, in all conditions, subjects evaluated the target, using the same four items as in Study 1.

To illustrate these changes, we present here the full text for the scenario about performance-enhancing drugs. Again, Brian is the target character, and Sam is the other person.

> Imagine that you are an athlete on a track team. Recently, your coach has become concerned that members of the team are using an illegal performance-enhancing drug called Vitronil. Vitronil use threatens your team's eligibility to compete, and gives individual athletes unfair advantages.
>
> Two of your teammates are named **Brian** and **Sam**. You know nothing about if either of them use Vitronil.

After this point, the passage differed across conditions, as follows:

> *Target-signals/condemnation condition:* One day, you are having a conversation with **Brian**. The two of you are discussing how different members of your team compete at meets. Specifically, you are talking about who stays clean, and who takes Vitronil. In your discussion, **Brian** emphasizes that he disapproves of taking Vitronil.
>
> *Other-signals/condemnation condition:* One day, you are having a conversation with **Sam**. The two of you are discussing how different members of your team compete at meets. Specifically, you are talking about who stays clean, and who takes Vitronil. In your discussion, **Sam** emphasizes that he disapproves of taking Vitronil.
>
> *Target-signals/direct-statement condition:* One day, you are having a conversation with **Brian**. The two of you are discussing how different members of your team compete at meets. Specifically, you are talking about who stays clean, and who takes Vitronil. In your discussion, **Brian** emphasizes that he does not take Vitronil.
>
> *Other-signals/direct-statement condition:* One day, you are having a conversation with **Sam**. The two of you are discussing how different members of your team compete at meets. Specifically, you are talking about who stays clean, and who takes Vitronil. In your discussion, **Sam** emphasizes that he does not take Vitronil.

After reading each vignette, subjects answered one comprehension question to ensure that they understood the story. They were allowed to continue participating even if they answered a question incorrectly; however, as in Study 1, we analyzed only the responses of subjects who showed perfect comprehension across all the vignettes. Overall, 99.1% of the subjects met this criterion (i.e., 7 subjects were excluded because of imperfect comprehension). We found no significant differences between conditions in the percentage of subjects who showed perfect comprehension, $\chi^2(3, N = 810) = 0.50$, $p = .919$. As in Study 1, we found high interitem reliability among our four individual dependent measures of evaluation of the targets ($\alpha = .88$), and we averaged these ratings to create a composite scale (see the Supplemental Material for analyses of the individual dependent measures).

Results

To test our predictions, we conducted a 2 (signaler: target vs. other person) × 2 (signal type: direct statement vs. condemnation) ANOVA predicting mean positive evaluations

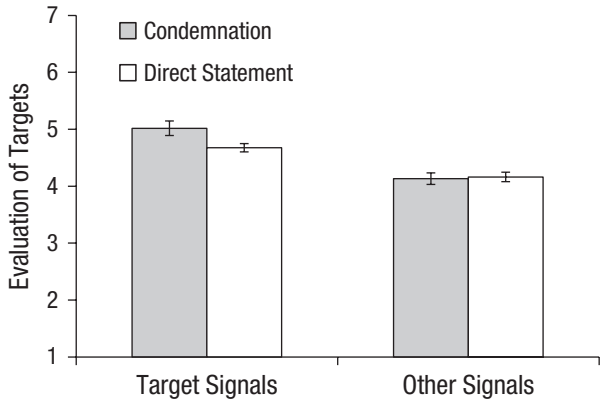


Fig. 2. Results from Study 2: mean composite evaluation of the targets as a function of the signaler and signal type. Error bars represent 95% confidence intervals.

of the targets across the vignettes (see Fig. 2). We found a significant main effect of signal type, $F(1, 799) = 9.92$, $p = .002$, $\eta_p^2 = .012$; subjects evaluated targets more positively in the condemnation condition ($M = 4.56$, $SD = 0.86$) than in the direct-statement condition ($M = 4.42$, $SD = 0.71$). This result demonstrates that, overall, reading about a conversation in which a transgression was morally condemned led subjects to evaluate the target more positively than reading about a conversation in which a character directly stated that he or she did not engage in that behavior.

We also found a significant main effect of signaler, $F(1, 799) = 198.62$, $p < .001$, $\eta_p^2 = .199$; subjects evaluated targets more positively when the target signaled ($M = 4.85$, $SD = 0.84$) than when the other person signaled ($M = 4.15$, $SD = 0.56$). This result demonstrates that, overall, targets' verbal signals of their moral goodness (condemnation and direct statements) successfully conferred reputational benefits to the targets.

Finally, as predicted, we observed a significant interaction of signaler and signal type, $F(1, 799) = 14.01$, $p < .001$, $\eta_p^2 = .017$; reading about condemnation of a transgression rather than a direct statement about behaving morally had a larger positive effect on evaluations of the target when the target signaled than when the other person signaled. In the target-signals condition, we observed a significant simple effect of signal type, with subjects evaluating the targets more positively when they engaged in condemnation ($M = 5.02$, $SD = 0.90$) than when they gave direct statements ($M = 4.68$, $SD = 0.74$), mean difference = 0.34, 95% CI = [0.18, 0.50], $t(397) = 4.15$, $p < .001$, $d = 0.42$. However, in the other-signals condition, we found no significant difference between the condemnation condition ($M = 4.13$, $SD = 0.53$) and the direct-statement condition ($M = 4.16$, $SD = 0.59$), mean difference = −0.03, 95% CI = [−0.14, 0.08], $t(402) = -0.53$, $p = .596$, $d = -0.05$. This result demonstrates that condemnation of a transgression can act as a stronger signal of one's own moral goodness than a direct statement of moral behavior.

Study 3

Our results in Study 2 suggest that condemnation can be a more persuasive signal of morality than a direct statement that one behaves morally. This implies that hypocrites may be judged even more negatively than straightforward liars: Their dishonesty may be more misleading, and may earn them larger undue reputation benefits. Additionally, hypocrites' false signals may be more destructive than liars' false statements (e.g., because their moral condemnation can malign and shame other people). In Study 3, we tested the prediction that hypocrites, who condemn transgressions they engage in, are judged more negatively than both (a) control transgressors, who engage in identical transgressions but do not condemn them, and (b) direct liars, who engage in identical transgressions but directly state that they do not.

Method

To test these predictions, we designed a new paradigm to evaluate perceptions of hypocrites. We created vignettes in which a target character discusses an acquaintance's moral transgression and then privately goes on to engage in the same transgression. We manipulated whether, in addition to committing the transgression, the target was a hypocrite, a (direct) liar, or neither.

On the basis of pilot testing, we selected transgressions that were perceived as more mild than those used in our first two studies (e.g., illegally downloading music, rather than using performance-enhancing drugs) to avoid floor effects—that is, to prevent the wrongness of the transgressions themselves from dominating subjects' evaluations of the targets (and consequently making it difficult to detect an effect of hypocrisy or lying). Additionally, we simplified our design by eliminating gender matching of the vignette characters and subjects and by dropping the comprehension questions. Finally, we adjusted our dependent measures. In our first two studies, we were interested in the signals that condemnation sends, so our dependent measures focused on predictions of the target's future moral behavior. In contrast, in Study 3 (and in Studies 4 and 5), we were interested in the implications of our theory for disapproval of hypocrites, so our dependent measures focused more on evaluations of the target as a person.

Design. We again presented subjects with vignettes and asked them to evaluate the target characters in the vignettes. In a three-condition, between-subjects design, we manipulated whether, before engaging in the relevant moral violation, the target character (a) condemned the

violation (*hypocrisy* condition), (b) directly stated that he or she did not engage in the violation (*liar* condition), or (c) said nothing (*control-transgressor* condition). We predicted that subjects would evaluate hypocrites as both worse than control transgressors and worse than liars.

Subjects. We recruited subjects online using MTurk. Because we were no longer predicting an interaction between conditions, we reduced our target cell size from 200 to 150 subjects, as per our standard protocol; thus, we precommitted to recruiting 450 subjects. A total of 461 people actually completed the survey. We analyzed the data of all subjects who had unique IP addresses and who had evaluated all the target characters. Our final sample consisted of 451 subjects (mean age = 35 years, 47% male).

Procedure. Each vignette described a conversation between two characters: a *target* (whom subjects would later evaluate) and a *friend* (whom subjects would not evaluate). In all conditions, this conversation began with the target and the friend discussing a mutual acquaintance. In this discussion, the friend mentioned that the mutual acquaintance often engaged in a particular moral transgression.

In the hypocrisy condition, the target responded to the friend by condemning the transgression. In contrast, in the liar condition, the target responded by directly stating that he or she did not engage in the relevant transgression. Finally, in the control-transgressor condition, we did not include any information about a response from the target. Shortly after this conversation ended, in all conditions, the target went on to commit the relevant violation.

For example, here is the full text for a scenario about downloading music illegally. In this scenario, Becky is the target character, and Amanda is the friend. In all conditions, the vignette began as follows:

> **Becky** and her friend **Amanda** are discussing a mutual acquaintance. **Amanda** mentions that the acquaintance often downloads music illegally from the Internet.

In the hypocrisy condition, the scenario continued,

> **Becky** says that she thinks it is morally wrong to download music illegally from the Internet. Shortly after their conversation, **Becky** goes online, and downloads music illegally.

In the liar condition, the scenario instead continued,

> **Becky** says that she doesn't download music illegally from the Internet. Shortly after their conversation, **Becky** goes online, and downloads music illegally.

Finally, in the control-transgressor condition, nothing was said about Becky's opinion or behavior, and the scenario simply ended with

> Shortly after their conversation, **Becky** goes online, and downloads music illegally.

Each subject was presented with four vignettes (in random order), about downloading music illegally, evading jury duty, ignoring phone calls from one's mother, and wasting paper by printing documents single-sided.

After reading each vignette, subjects evaluated the target. We asked subjects to rate how good a person the target was, how much they liked the target, how honest they thought the target was, and how trustworthy they thought the target was. Further, as a manipulation check, and to conduct an exploratory investigation into how subjects conceptualize "hypocrisy," we asked them to rate how hypocritical the target was. These five dependent measures were presented in random order for each vignette. Subjects responded to each item on a sliding scale, with anchors reading *not at all [trait]* to *very [trait]* (e.g., *not at all trustworthy* to *very trustworthy*). The sliding scales did not have any numerical labels, but responses were translated to scores ranging from 0 (*not at all*) to 100 (*very*).

We found high interitem reliability among our four primary individual dependent measures (i.e., excluding our hypocrisy variable, which was used as a manipulation check; $\alpha = .96$). Thus, as in Studies 1 and 2, we averaged these ratings to create a composite scale representing positive evaluations of the target and used this composite as our dependent variable in the analyses reported here (analyses investigating the individual dependent measures are reported in the Supplemental Material). We note that including the hypocrisy measure in our composite variable did not qualitatively affect our conclusions; however, results from the hypocrisy measure followed a somewhat different pattern than results from the other measures, and provide insight into the ways that subjects use the term "hypocrisy" (see the Supplemental Material).

Results

To test our prediction, we conducted a one-way ANOVA investigating the effect of condition on positive evaluations of the targets across the vignettes (see Fig. 3). We found a significant effect of condition, $F(2, 448) = 26.48$, $p < .001$, $\eta_p^2 = .106$. We followed up with pairwise comparisons and found that hypocrites ($M = 32.07$, $SD = 15.93$) were evaluated as worse than liars ($M = 36.15$, $SD = 15.37$),

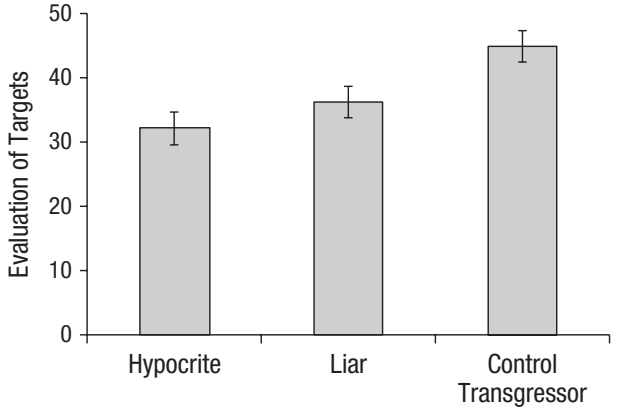


Fig. 3. Results from Study 3: mean composite evaluation of the targets as a function of condition. Error bars represent 95% confidence intervals.

mean difference = −4.08, 95% CI = [−7.63, −0.53], $t(299) = -2.26$, $p = .025$, $d = -0.26$, who were evaluated as worse than control transgressors ($M = 44.77$, $SD = 14.89$), mean difference = −8.62, 95% CI = [−12.04, −5.20], $t(301) = -4.96$, $p < .001$, $d = -0.57$. This result confirmed our prediction that hypocrites would be seen as worse than liars. It also demonstrates that people disapprove of false signalers (hypocrites and liars) more than they disapprove of people who commit the same transgressions but do not condemn other people for those transgressions or lie about engaging in them.

Study 4

In Study 4, we moved to testing our theory's key prediction that hypocrites are perceived negatively *because* of their false signals. If the negative perception of hypocrisy is caused by hypocrites' false signaling, people we refer to as *honest hypocrites* (who avoid sending false signals) should not be judged negatively. Honest hypocrites fail to live up to their own moral standards and criticize other people for behaviors they themselves engage in, but admit that they sometimes commit the deeds they condemn. Honest hypocrites thus condemn without signaling, and analyzing subjects' evaluations of such characters allowed us to test our false-signaling theory against several alternatives: If traditional hypocrites are disliked because they are inconsistent, unpredictable, weak willed, or intentionally immoral, people should dislike honest hypocrites, too. However, if people dislike traditional hypocrites because they send false signals, honest hypocrites should not be judged as worse than nonhypocritical transgressors.

Method

Design. In Study 4, we compared evaluations of honest hypocrites with evaluations of traditional hypocrites and control transgressors. To this end, we used the same design as in Study 3, but replaced the liar condition with an honest-hypocrite condition. In the honest-hypocrite condition, the target responded to the friend by stating that he or she believed the behavior in question to be morally wrong but sometimes behaved that way anyway.

Thus, in a three-condition, between-subjects design, we manipulated whether, before engaging in the relevant moral transgression, the target (a) condemned the violation (traditional-hypocrite condition), (b) condemned the violation but explicitly negated any signaling value of the condemnation (honest-hypocrite condition), or (c) said nothing (control-transgressor condition). We predicted that subjects would evaluate honest hypocrites as no worse than control transgressors (and that traditional hypocrites would be seen as worse than both honest hypocrites and control transgressors).

Subjects. As in Study 3, we recruited subjects online using MTurk. We precommitted to recruiting 450 subjects ($n = 150$ per condition), and a total of 457 actually completed the survey. We analyzed responses from all subjects who had unique IP addresses and had evaluated all the vignettes. Our final sample consisted of 452 subjects (mean age = 35 years, 41% male).

Procedure. The procedure for presenting the vignettes and measuring evaluations of targets was identical to that in Study 3, except that we replaced the liar condition with an honest-hypocrite condition. For example, here is the full text for the scenario about downloading music illegally. Again, Becky is the target character, and Amanda is the friend. The scenario began as follows:

> **Becky** and her friend **Amanda** are discussing a mutual acquaintance. **Amanda** mentions that the acquaintance often downloads music illegally from the Internet.

In the traditional-hypocrite condition, the scenario continued with

> **Becky** says that she thinks it is morally wrong to download music illegally from the Internet. Shortly after their conversation, **Becky** goes online, and downloads music illegally.

In the honest-hypocrite condition, the passage instead continued with

> **Becky** says that she thinks it is morally wrong to download music illegally from the Internet, but that she sometimes does it anyway. Shortly after their conversation, **Becky** goes online, and downloads music illegally.

Finally, in the control-transgressor condition, nothing was said about Becky's opinion or behavior, and the passage simply ended with

> Shortly after their conversation, **Becky** goes online, and downloads music illegally.

As in Study 3, we found a high interitem reliability among our four individual primary dependent measures (α = .94) and averaged responses to them to create a single composite measure (see the Supplemental Material for analyses investigating the individual dependent measures, including the measure of subjects' concept of "hypocrisy").

Results

To test our prediction, we conducted a one-way ANOVA investigating the effect of condition on positive evaluations of the targets across the vignettes (see Fig. 4). We found a significant effect of condition, $F(2, 449) = 35.62$, $p < .001$, $\eta_p^2 = .137$. Pairwise comparisons revealed that traditional hypocrites ($M = 34.35$, $SD = 13.76$) were evaluated more negatively than both honest hypocrites ($M = 46.62$, $SD = 14.01$), mean difference = −12.28, 95% CI = [−15.41, −9.14], $t(301) = -7.69$, $p < .001$, $d = -0.88$, and control transgressors ($M = 45.21$, $SD = 13.84$), mean difference = −10.87, 95% CI = [−13.99, −7.74], $t(300) = -6.84$, $p < .001$, $d = -0.79$. Additionally, we found that evaluations of honest hypocrites and control transgressors did not differ, mean difference = 1.41, 95% CI = [−1.76, 4.58], $t(297) = 0.88$, $p = .382$, $d = 0.10$. Thus, as predicted, our honest-hypocrite manipulation fully eliminated disapproval of hypocrisy; honest hypocrites received ratings that were no worse than the ratings of control transgressors.

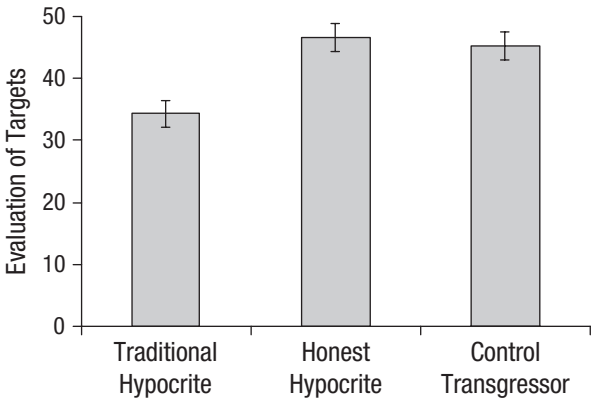


Fig. 4. Results from Study 4: mean composite evaluation of the targets as a function of condition. Error bars represent 95% confidence intervals.

Study 5

In Study 4, honest hypocrites, whose condemnation was stripped of its signaling function, were not judged more negatively than control transgressors, who engaged in transgressions without condemnation. However, it is possible that honest hypocrites are in fact judged negatively for their hypocrisy, but are given additional credit for voluntarily disclosing their transgressions, which offsets the negative evaluation of their hypocrisy. In Study 5, we tested this alternative explanation by investigating evaluations of hypocrites who disclosed transgressions that were unrelated to their condemnation, and thus did not negate the false signals implied by their condemnation.

Method

Design. In Study 5, we modified our Study 4 design to include a *disclosure-hypocrite* condition that involved hypocrisy (condemnation followed by transgression) and a disclosure about a transgression unrelated to the condemnation. To this end, we altered our vignettes so that the targets in all conditions committed two moral transgressions, rather than one, and we presented our vignettes in a way that naturally introduced these two transgressions. Then, in a four-condition, between-subjects design, we manipulated whether, before engaging in these two transgressions, the targets (a) condemned one transgression (traditional hypocrite), (b) condemned one transgression and admitted to engaging in the other transgression (disclosure hypocrite), (c) condemned one transgression and admitted to engaging in that same transgression (honest hypocrite), or (d) said nothing (control transgressor).

Thus, both the disclosure hypocrite and the honest hypocrite ultimately committed the same two violations. They also each condemned one of the violations and admitted to committing one of the violations. However, only the honest hypocrite admitted to committing the same violation he or she condemned, and thus negated the false signal implied by that condemnation. We predicted that whereas honest hypocrites would be seen as no worse than control transgressors (because their hypocrisy did not involve false signaling), disclosure hypocrites would be seen as no better than traditional hypocrites (because their hypocrisy still involved false signaling, despite also involving disclosure).

Subjects. As in Study 4, we recruited subjects online using MTurk. We precommitted to recruiting 150 subjects per condition (i.e., a total of 600 subjects). A total of 612 subjects actually completed the survey. All of these subjects had unique IP addresses and had evaluated all the vignettes, so none were excluded from analyses. Thus, our final sample consisted of 612 subjects (mean age = 34 years, 48% male).

Procedure. To implement our design, we collapsed our four vignettes (in which each target committed one violation) into two vignettes (in which each target committed two violations). Specifically, in one vignette, the target downloaded music illegally and ignored his or her mother's phone calls, and in the other, the target tried to get out of jury duty and wasted paper by printing single-sided. We presented subjects with both vignettes in a random order.

We modified each vignette to structure the target's conversation around the two moral transgressions at hand. Specifically, we introduced the two moral issues by explaining that the target and the friend were discussing issues in their lives and then listing the relevant topics as examples (e.g., downloading music illegally and answering parents' phone calls). Then, depending on the condition, the vignette presented any relevant condemnation and disclosure information. Finally, the vignette indicated that the target went on to commit both transgressions.

For example, here is the full text for the scenario about downloading music illegally and answering parents' phone calls. Again, Becky is the target character, and Amanda is the friend. In all conditions, the vignette began as follows:

> **Becky** and her friend **Amanda** are discussing issues in their lives, like downloading music and answering their parents' phone calls.

In the traditional-hypocrite condition, the vignette continued,

> **Becky** tells **Amanda** that she thinks it is morally wrong when people download music illegally from the Internet. Shortly after their conversation, **Becky** goes online, and downloads music illegally. She also notices that her mother is calling, and ignores the call.

In the disclosure-hypocrite condition, the passage instead read,

> **Becky** tells **Amanda** that she thinks it is morally wrong when people download music illegally from the Internet, but that she sometimes ignores her mother's phone calls. Shortly after their conversation, **Becky** goes online, and downloads music illegally. She also notices that her mother is calling, and ignores the call.

In the honest-hypocrite condition, the passage read,

> **Becky** tells **Amanda** that she thinks it is morally wrong when people download music illegally from the Internet, but that she sometimes does it anyway. Shortly after their conversation, **Becky** goes online, and downloads music illegally. She also notices that her mother is calling, and ignores the call.

Finally, in the control-transgressor condition, there was no mention of Becky's opinion or behavior, and the passage simply ended with

> Shortly after their conversation, **Becky** goes online, and downloads music illegally. She also notices that her mother is calling, and ignores the call.

We orthogonally counterbalanced both (a) which of the two transgressions we listed first (when introducing them and explaining that the target engaged in them) and (b) which of the two transgressions was condemned by the targets in the three hypocrite conditions (and consequently, which transgression was disclosed by the targets in the disclosure-hypocrite condition—as this was always the noncondemned transgression).

With the exception of these modifications, Study 5 was identical to Study 4, and thus used the same dependent measures. We again found a high interitem reliability among our four individual primary dependent measures ($\alpha = .93$) and averaged them to create a single composite measure (see the Supplemental Material for analyses investigating the individual dependent measures, including the measure of subject's concept of "hypocrisy").

Results

To test our predictions, we conducted a one-way ANOVA investigating the effect of condition on positive evaluations of the targets across the vignettes (see Fig. 5). We found a significant effect of condition, $F(3, 608) = 28.60$, $p < .001$, $\eta_p^2 = .124$. As in Study 4, pairwise comparisons revealed that honest hypocrites ($M = 41.94$, $SD = 14.51$) were evaluated more positively than traditional hypocrites ($M = 28.22$, $SD = 15.30$), mean difference = 13.73, 95% CI = [10.38, 17.08], $t(305) = -8.06$, $p < .001$, $d = -0.92$, and were seen as no worse than control transgressors ($M = 38.85$, $SD = 15.94$). In fact, honest hypocrites were seen as marginally better than control transgressors in Study 5, mean difference = 3.09, 95% CI = [−0.34, 6.53], $t(303) = 1.77$, $p = .077$, $d = 0.20$.

Critically, pairwise comparisons also revealed that disclosure hypocrites, who merely admitted to committing a moral transgression but did not negate the false signal implied by their condemnation, did not receive similarly positive evaluations. Disclosure hypocrites ($M = 30.03$, $SD = 15.95$) were seen as significantly worse than honest hypocrites, mean difference = −11.91, 95% CI = [−15.34, −8.48], $t(304) = -6.83$, $p < .001$, $d = -0.78$, and were not seen as

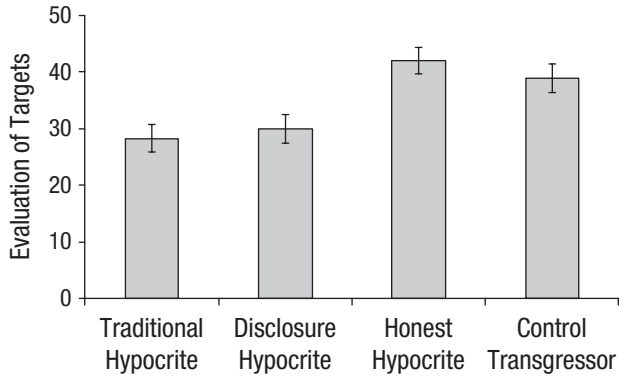


Fig. 5. Results from Study 5: mean composite evaluation of the targets as a function of condition. Error bars represent 95% confidence intervals.

significantly better than traditional hypocrites, mean difference = 1.81, 95% CI = [−1.70, 5.32], $t(305) = 1.02$, $p = .310$, $d = 0.12$. This result demonstrates that, as predicted, mere disclosure is insufficient to eliminate subjects' disapproval of hypocrites: Hypocritical targets must use disclosure that negates the false signals implied by their condemnation in order to eliminate this disapproval.

General Discussion

We have sought to explain why hypocrites—who condemn transgressions they engage in—are seen as worse than individuals who commit the same transgressions without condemning them. The puzzle, as we see it, is this: Condemnation of bad behavior is typically seen as virtuous (because it discourages that behavior), and people who do not condemn bad behavior can be seen as second-order free riders (Yamagishi, 1986). So why do hypocrites get moral blame—not credit—for their condemnation?

Our experiments provide an answer: Hypocrites are disliked because they falsely signal that they behave morally. This theory explains that hypocrites do in fact free-ride; they do so not by refusing to condemn bad behavior, but by using condemnation to imply that they will behave morally—without incurring the costs of actually doing so.

Our findings, by elucidating the conditions under which hypocrisy is perceived negatively by other people, may shed light on previous work showing that hypocrites themselves experience hypocrisy as aversive, and may explain why a fear of hypocrisy in public contexts is especially effective at motivating virtuous behavior (Aronson et al., 1991).

Our results support our theory of hypocrisy by demonstrating that condemnation of immoral behavior is perceived as a signal of moral behavior (Study 1) that can be more convincing than directly stating that one behaves morally (Study 2). These results are consistent with theories that moral language conveys much beyond its literal meaning (Strandberg, 2012), and that condemnation communicates information about one's values and behaviors (Baumeister et al., 2004). Condemnation may also be interpreted as a more sincere signal of morality compared with direct statements because, at first blush, condemning other people is less obviously self-promotional than stating that one behaves morally. Because people actively monitor social information for its veracity (Barasch, Levine, Berman, & Small, 2014; Fein, Hilton, & Miller, 1990; Hess & Hagen, 2006; Lin-Healy & Small, 2012), and overt self-promotion can thus backfire (Gordon, 1996), condemnation may be a more persuasive signal.

These results also build on the finding that people who punish selfishness in economic-game experiments are trusted not to act selfishly themselves (Barclay, 2006; Horita, 2010; Jordan et al., 2016; Nelissen, 2008; Raihani & Bshary, 2015a, 2015b). Verbal condemnation can function as costly punishment (Fehr & Fischbacher, 2004): It harms the transgressor's reputation and is also risky for the condemner—because the transgressor might retaliate (e.g., by publicizing the condemner's misdeeds). Previous research has shown that punishment can function as a costly signal (Zahavi, 1975) of morality, so long as punishing is less costly for people who typically behave morally than for those who behave immorally (Jordan et al., 2016; Jordan & Rand, 2016). Perhaps, then, verbal condemnation of immoral behavior is perceived as a strong signal because it acts as a costly signal—in ways that direct statements that one behaves morally (which carry few costs) do not. Future research should explicitly investigate the effect of cost on perceptions of condemnation of immorality, direct statements of one's own morality, and other signals of morality. Moreover, future research should investigate whether praising *good* behavior signals morality. Praising an action one does not engage in may inspire less outrage than condemning behavior one does engage in, if praise serves as a weaker signal of morality because it is less costly (i.e., the target of praise is unlikely to retaliate).

Studies 1 and 2 also extend theories about *credibility-enhancing displays* (Henrich, 2009)—costly indicators that one holds a particular belief (e.g., eating a mushroom to signal that one believes it is healthy). Much as actions can undermine credibility, hypocrisy negates a signal implied by condemnation. Whereas credibility-enhancing displays signal beliefs about states of the world, we have shown that condemnation signals future moral behavior.

Our theory is further supported by Studies 3 through 5, which showed that people dislike hypocrites more than direct liars, and that this is because hypocrites falsely signal. One straightforward explanation for why hypocrites' false signals inspire moral outrage is that misleading other

people is generally regarded as wrong (Bell & Whaley, 1991)—and hypocrites are especially misleading, because condemnation is an especially convincing signal.

A hypocrite's false signals may rouse further disapproval, moreover, because they lead to negative outcomes, such as unfairly boosting the hypocrite's reputation or shaming other people into changing their behavior while the hypocrite carries on. Furthermore, unlike direct statements that one behaves morally, condemnation can harm other people by maligning the condemned—which may make hypocrisy seem particularly wrong (Crockett, Kurth-Nelson, Siegel, Dayan, & Dolan, 2014; Eber, 2007; Smith, Parrott, Ozer, & Moniz, 1994). Consistent with the hypothesis that hypocrites are judged as worse than liars for reasons beyond their being more misleading, our supplementary analyses showed that hypocrites were rated especially negatively relative to liars on measures of being likeable and a good person; the difference between ratings of hypocrites and liars was smaller on measures of trust and honesty (see the Supplemental Material).

An important future direction is to investigate perceptions of condemnation and hypocrisy across cultures. Our data are limited to American MTurk samples, which raises questions about the generalizability of our results (Henrich, Heine, & Norenzayan, 2010). (We note, though, that within our samples, results were robust across demographic variables; see the Supplemental Material.) Although condemnation appears to be widespread across cultures, its prevalence does vary substantially (Henrich et al., 2006). How does this variance correlate with the signal value of condemnation and with disapproval of hypocrites? Future research should address this question, and also investigate hypocrisy in less contrived contexts (e.g., reported examples from daily life) and other culturally relevant domains (e.g., religious hypocrisy).

In conclusion, we propose that hypocrites are disliked because their condemnation falsely signals moral goodness. We have supported this theory with evidence that when condemnation's signaling value is negated, hypocrisy is forgiven.

Action Editor

Leaf Van Boven served as action editor for this article.

Author Contributions

J. J. Jordan, R. Sommers, P. Bloom, and D. G. Rand developed the study concept. J. J. Jordan, P. Bloom, and D. G. Rand designed Studies 1 and 2. J. J. Jordan and R. Sommers designed Studies 3 through 5. J. J. Jordan collected data and performed the data analysis. All four authors wrote the manuscript.

Acknowledgments

We gratefully acknowledge helpful comments and assistance from Adam Bear, Adam Chekroud, Fabian Schellhaas, and Amanda Zheutlin.

Declaration of Conflicting Interests

The authors declared that they had no conflicts of interest with respect to their authorship or the publication of this article.

Funding

This research was funded by a grant from the John Templeton Foundation.

Supplemental Material

Additional supporting information can be found at http://journals.sagepub.com/doi/suppl/10.1177/0956797616685771

Open Practices

All data and materials have been made publicly available via the Open Science Framework and can be accessed at https://osf.io/pszjz/. The complete Open Practices Disclosure for this article can be found at http://journals.sagepub.com/doi/suppl/10.1177/0956797616685771. This article has received badges for Open Data and Open Materials. More information about the Open Practices badges can be found at http://www.psychologicalscience.org/publications/badges.

References

Alicke, M., Gordon, E., & Rose, D. (2013). Hypocrisy: What counts? *Philosophical Psychology*, *26*, 673–701.

Aronson, E., Fried, C., & Stone, J. (1991). Overcoming denial and increasing the intention to use condoms through the induction of hypocrisy. *American Journal of Public Health*, *81*, 1636–1638.

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

Barclay, P. (2006). Reputational benefits for altruistic punishment. *Evolution & Human Behavior*, *27*, 325–344.

Barden, J., Rucker, D. D., & Petty, R. E. (2005). "Saying one thing and doing another": Examining the impact of event order on hypocrisy judgments of others. *Personality and Social Psychology Bulletin*, *31*, 1463–1474.

Batson, C. D., Thompson, E. R., Seuferling, G., Whitney, H., & Strongman, J. A. (1999). Moral hypocrisy: Appearing moral to oneself without being so. *Journal of Personality and Social Psychology*, *77*, 525–537.

Baumeister, R. F., Zhang, L., & Vohs, K. D. (2004). Gossip as cultural learning. *Review of General Psychology*, *8*, 111–121.

Bell, J. B., & Whaley, B. (1991). *Cheating and deception*. New Brunswick, NJ: Transaction.

Berkowitz, L., & Walker, N. (1967). Laws and moral judgments. *Sociometry*, *30*, 410–422.

Crockett, M. J., Kurth-Nelson, Z., Siegel, J. Z., Dayan, P., & Dolan, R. J. (2014). Harm to others outweighs harm to self in moral decision making. *Proceedings of the National Academy of Sciences, USA*, *111*, 17320–17325.

Cushman, F., Young, L., & Hauser, M. (2006). The role of conscious reasoning and intuition in moral judgment: Testing

three principles of harm. *Psychological Science*, *17*, 1082–1089.

Eber, N. (2007). The performance-enhancing drug game reconsidered: A fair play approach. *Journal of Sports Economics*, *9*, 318–327.

Fehr, E., & Fischbacher, U. (2004). Third-party punishment and social norms. *Evolution & Human Behavior*, *25*, 63–87.

Fein, S., Hilton, J. L., & Miller, D. T. (1990). Suspicion of ulterior motivation and the correspondence bias. *Journal of Personality and Social Psychology*, *58*, 753–764.

Feinberg, M., Willer, R., & Schultz, M. (2014). Gossip and ostracism promote cooperation in groups. *Psychological Science*, *25*, 656–664.

Feinberg, M., Willer, R., Stellar, J., & Keltner, D. (2012). The virtues of gossip: Reputational information sharing as prosocial behavior. *Journal of Personality and Social Psychology*, *102*, 1015–1030.

Gordon, R. A. (1996). Impact of ingratiation on judgments and evaluations: A meta-analytic investigation. *Journal of Personality and Social Psychology*, *71*, 54–70.

Henrich, J. (2009). The evolution of costly displays, cooperation and religion: Credibility enhancing displays and their implications for cultural evolution. *Evolution & Human Behavior*, *30*, 244–260.

Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world? *Behavioral & Brain Sciences*, *33*, 61–83.

Henrich, J., McElreath, R., Barr, A., Ensminger, J., Barrett, C., Bolyanatz, A., . . . Ziker, J. (2006). Costly punishment across human societies. *Science*, *312*, 1767–1770.

Hess, N. H., & Hagen, E. H. (2006). Psychological adaptations for assessing gossip veracity. *Human Nature*, *17*, 337–354.

Horita, Y. (2010). Punishers may be chosen as providers but not as recipients. *Letters on Evolutionary Behavioral Science*, *1*(1), 6–9.

Jordan, J. J., Hoffman, M., Bloom, P., & Rand, D. G. (2016). Third-party punishment as a costly signal of trustworthiness. *Nature*, *530*, 473–476.

Jordan, J. J., & Rand, D. G. (2016). *Building costly signaling from the ground up: A model of third-party punishment as a costly signal of exposure to repeated interactions*. Retrieved from https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2794084

Lin-Healy, F., & Small, D. A. (2012). Cheapened altruism: Discounting personally affected prosocial actors. *Organizational Behavior and Human Decision Processes*, *117*, 269–274.

McKinnon, C. (1991). Hypocrisy, with a note on integrity. *American Philosophical Quarterly*, *28*, 321–330.

Nelissen, R. M. A. (2008). The price you pay: Cost-dependent reputation effects of altruistic punishment. *Evolution & Human Behavior*, *29*, 242–248.

Raihani, N. J., & Bshary, R. (2015a). The reputation of punishers. *Trends in Ecology & Evolution*, *30*, 98–103.

Raihani, N. J., & Bshary, R. (2015b). Third-party punishers are rewarded, but third-party helpers even more so. *Evolution*, *69*, 993–1003.

Righetti, F., & Finkenauer, C. (2011). If you are able to control yourself, I will trust you: The role of perceived self-control in interpersonal trust. *Journal of Personality and Social Psychology*, *100*, 874–886.

Shklar, J. N. (1984). *Ordinary vices*. Cambridge, MA: Harvard University Press.

Smith, R. H., Parrott, W. G., Ozer, D., & Moniz, A. (1994). Subjective injustice and inferiority as predictors of hostile and depressive feelings in envy. *Personality and Social Psychology Bulletin*, *20*, 705–711.

Strandberg, C. (2012). A dual aspect account of moral language. *Philosophy and Phenomenological Research*, *84*, 87–122.

Tedeschi, J. T., Schlenker, B. R., & Bonoma, T. V. (1971). Cognitive dissonance: Private ratiocination or public spectacle? *American Psychologist*, *26*, 685–695.

Williams, K. D., Forgas, J. P., & von Hippel, W. (Eds.). (2005). *The social outcast: Ostracism, social exclusion, rejection, and bullying*. New York, NY: Psychology Press.

Yamagishi, T. (1986). The provision of a sanctioning system as a public good. *Journal of Personality and Social Psychology*, *51*, 110–116.

Zahavi, A. (1975). Mate selection—a selection for a handicap. *Journal of Theoretical Biology*, *53*, 205–214.

DOI: 10.1177/0956797616685771

Supplementary information
for
Why do we hate hypocrites? Evidence for a theory of false signaling

Jillian Jordan, Roseanna Sommers, Paul Bloom, and David Rand

DOI: 10.1177/0956797616685771

Supplementary analyses

Analyses of "hypocrisy" DV in Studies 3-5

In Studies 3-5, we included hypocrisy as a dependent variable (DV), both as a manipulation check, and as a way to investigate subjects' conceptions of the term "hypocrisy." Including the hypocrisy measure in our positive evaluation composite does not qualitatively affect our conclusions; however, results from the hypocrisy measure do pattern somewhat differently than the other items, and provide insight into the ways that subjects use the term "hypocrisy." Thus, we report these analyses here.

Study 3

In Study 3, we compared traditional hypocrites to liars and control transgressors. Here we report hypocrisy ratings across these conditions.

For this analysis, we conducted a one-way ANOVA investigating the effect of condition on hypocrisy ratings. We found a significant effect of condition, $F(2,448) = 71.49$, $p < .001$, $\eta_p^2 = .242$. We followed up with pairwise comparisons, and found that hypocrites ($M = 76.87$, $SD = 22.72$) were seen as more hypocritical than liars ($M = 70.38$, $SD = 19.06$) ($t(299) = 2.69$, $p = .008$, $d = .31$), who in turn were seen as more hypocritical than control transgressors ($M = 49.24$, $SD = 20.83$) ($t(301) = 9.22$, $p < .001$, $d = 1.06$).

Thus, unsurprisingly, hypocrites were perceived as more hypocritical than non-hypocrites (liars and control transgressors); this serves as a manipulation check. Interestingly, we found that non-hypocrites were far from the floor on our hypocrisy scale: control transgressors were rated around the midpoint, and liars were rated almost three quarters of the way to the maximum rating of hypocrisy. Furthermore, we found that liars were seen as more hypocritical than control transgressors. These results suggest that (at least some) subjects applied the term "hypocritical"

loosely, perhaps as a way to condemn any form of negative or deceptive behavior—rather than strictly defining hypocrites as those who both commit and condemn the same transgression.

Study 4

In Study 4, we compared traditional hypocrites to honest hypocrites and control transgressors. Here we report hypocrisy ratings across these conditions.

For this analysis, we again conducted a one-way ANOVA investigating the effect of condition on hypocrisy. We found a significant effect of condition, $F(2,449) = 83.19$, $p < .001$, $\eta_p^2 = .270$. We followed up with pairwise comparisons, and found that traditional hypocrites ($M = 78.46$, $SD = 19.65$) were seen as more hypocritical than honest hypocrites ($M = 64.92$, $SD = 20.19$) ($t(301) = 5.91$, $p < .001$, $d = .68$), who in turn were seen as more hypocritical than control transgressors ($M = 48.16$, $SD = 21.45$) ($t(297) = 6.96$, $p < .001$, $d = .80$).

Thus, we replicate our findings from Study 3, in which traditional hypocrites were perceived as more hypocritical than control transgressors, and control transgressors were near the midpoint of the scale. Furthermore, we find that honest hypocrites are perceived as intermediately hypocritical: they are seen as less hypocritical than traditional hypocrites, but as more hypocritical than control transgressors. This again suggests that subjects may use the term "hypocritical" loosely as a way to condemn negative and deceptive behavior. However, it also suggests that to some degree, subjects recognize the "hypocrisy" of honest hypocrites—even though they seem not to translate this judgment into negative evaluations. As Study 4 showed, evaluations of honest hypocrites were no more negative than evaluations of control transgressors.

Study 5

In Study 5, we compared traditional hypocrites to disclosure hypocrites, honest hypocrites, and control transgressors. Here we report hypocrisy ratings across these conditions.

For this analysis, we again conducted a one-way ANOVA investigating the effect of condition on hypocrisy. We found a significant effect of condition, $F(3,608) = 69.29$, $p < .001$, $\eta_p^2 = .255$. We followed up with pairwise comparisons, and found that traditional hypocrites ($M = 80.75$, $SD = 21.10$) were seen as more hypocritical than disclosure hypocrites ($M = 75.60$, $SD = 19.90$) ($t(305) = 2.20$, $p = .028$, $d = .25$), who were seen as more hypocritical than honest hypocrites ($M = 65.07$, $SD = 23.06$) ($t(304) = 4.28$, $p < .001$, $d = .49$), who were seen as more hypocritical than control transgressors ($M = 48.12$, $SD = 21.37$) ($t(303) = 6.66$, $p < .001$, $d = .76$).

Thus, we replicate our pattern of results from Study 4. Interestingly, we also find that disclosure hypocrites are seen as intermediately hypocritical, relative to traditional and honest hypocrites. Disclosure hypocrites' decision to disclose transgressions that are unrelated to their condemnation therefore appears to make them seem somewhat less hypocritical than traditional hypocrites, although this boost does not translate into significantly more positive evaluations of them.

Conclusions

In sum, subjects' use of the term "hypocritical" appears to reflect the general sense that a target behaves badly or deceptively: targets who were evaluated negatively were often called "hypocritical," even when they did not conform to the understanding of hypocrisy that we had in mind (condemning a transgression that one engages in). Nonetheless, subjects also seemed *somewhat* attuned to our understanding of hypocrisy: honest hypocrites (who did condemn a transgression they engaged in) were seen as more hypocritical than control transgressors (who did not) despite not being evaluated more negatively overall.

Thus, subjects displayed a mixed conception of hypocrisy. One interesting open question is whether this pattern reflects individual differences in how the term is used, or whether single

DOI: 10.1177/0956797616685771

individuals would show the same patterns we observed overall in a within-subjects design. Another interesting question is whether we would have obtained different results had our only DV been hypocrisy. Perhaps part of why hypocrisy patterned, to some extent, with general negative evaluations (and why non-hypocrites were rated as hypocritical) is that subjects were rating hypocrisy on the same page as they were providing their general evaluations of targets, and there may have been "spillover" between measures.

DOI: 10.1177/0956797616685771

Analyses of individual dependent measures in Studies 1-5

In all of our analyses presented in the main text, we used a composite positive evaluation measure as our dependent variable. In Studies 1 and 2, this composite was computed as the average of ratings of (i) the perceived likelihood that the target would not transgress in the future, (ii) specific trust of the target, (iii) general trust of the target, and (iv) liking of the target. In Studies 3-5, this composite was computed as the average of ratings of (i) how good a person the target was, (ii) liking of the target, (iii) trust of the target, and (iv) honesty of the target. In all five studies, we computed and analyzed a composite variable because we found high inter-item reliability among our four dependent variables. For completeness, we also investigated possible differences between dependent measures, and found that results were largely qualitatively equivalent across individual measures. Here, we report our analyses of individual dependent measures.

Specifically, we report two sets of analyses. First, for each of Studies 1-5, we conducted mixed-effects ANOVAs where condition variables are between-subject factors, and measure type (i.e. which DV?) is a within-subject factor. We found significant interactions between condition variables and measure type for each study. Second, based on these interactions, for each study, we redid our ANOVAs from the main text (with condition variables as between-subject factors, and no within-subject factors) separately for each measure type, and plotted the results by condition separately for each measure type. These separate ANOVAs and plots demonstrate that (i) the overall pattern of results by condition is robust and holds across dependent measures and (ii) the size of the differences between some pairs of conditions does vary between dependent measures (the implications of this variation is discussed below).

DOI: 10.1177/0956797616685771

Mixed-effects ANOVAs

We begin by reporting, for each study, whether the effects of condition variables differed significantly by dependent measure type. For Study 1, we conducted a mixed-effects ANOVA with evaluation of the target (averaged across the four vignettes) as the dependent variable. We used *condemnation* (target condemns vs. other condemns), *information* (good information vs. no information), and their interaction as between-subjects factors, and we used *measure type* (likelihood of future transgression vs. specific trust vs. general trust vs. liking) as a within-subjects factor. Crucially, we were interested in interactions between the between-subjects factors and the within-subjects factor, which answer the question of whether condition effects differed significantly by dependent measure type. We found a significant two-way interaction between condemnation and measure type, $F(3,1845) = 16.43$, $p < .001$, as well as a significant two-way interaction between information and measure type, $F(3,1845) = 5.61$, $p < .001$. We did not, however, find a significant three-way interaction between condemnation, information, and measure type, $F(3,1845) = 1.34$, $p = .261$. These results demonstrate that the main effects of condemnation and information both differ significantly by measure type, but the interaction between condemnation and information does not.

For Study 2, we repeated this analysis approach, with the substitution that our between-subjects condition variables were *signaling* (target signals vs. other signals) and *signal type* (direct statements vs. moral condemnation). We found a significant two-way interaction between signaling and measure type, $F(3,2397) = 7.30$, $p < .001$, a significant two-way interaction between signal type and measure type, $F(3,2397) = 3.80$, $p = .010$. We also found a significant three-way interaction between signaling, signal type, and measure type, $F(3,2397) = 2.86$, $p =$

.035. These results demonstrate that the main effects of signaling and signal type both differ significantly by measure type, as does the interaction between signaling and signal type.

For Study 3, we again conducted a mixed-effects ANOVA with evaluation of the target (averaged across the four vignettes) as the dependent variable. We used *condition* (hypocrite vs. liar vs. control transgressor) as a between-subjects factor, and we used *measure type* (good person vs. liking vs. trust vs. honesty) as a within-subjects factor. We were interested in the interaction between condition and measure type, and found a significant interaction, $F(6,1344) = 15.56$, $p <. 001$. This demonstrates that the effect of condition differs significantly by measure type.

For Study 4, we repeated this analysis approach, with the substitution that our condition variable tracked different conditions (traditional hypocrite vs. honest hypocrite vs. control transgressor). We found a significant interaction between condition and measure type, $F(6,1347) = 19.52$, $p < .001$, demonstrating that the effect of condition differs significantly by measure type.

For Study 5, we repeated this analysis approach, with the substitution that our condition variable tracked different conditions (traditional hypocrite vs. honest hypocrite vs. disclosure hypocrite vs. control transgressor). We found a significant interaction between condition and measure type, $F(9,1824) = 4.64$, $p < .001$, demonstrating that the effect of condition differs significantly by measure type.

DOI: 10.1177/0956797616685771

Analyses by measure type

Thus, in all five studies, we found evidence that the effects of our condition variables differed significantly by measure type. To investigate these differences, we redid our ANOVAs from the main text (with condition variables as between-subject factors, and no within-subject favors), and present follow-up pairwise *t*-tests, separately for each measure type. We also plotted results by condition separately for each measure type. To facilitate a comparison of results across measure type, for each study, we present statistical tests for each dependent measure in a table. For all studies, we find qualitatively similar patterns of results across measures, although effect sizes vary, and not all statistical tests for all measures reach significance. Below, these results by measure are presented for each study.

DOI: 10.1177/0956797616685771

The key Study 1 results, reported in the main text, were a significant positive Information X Condemnation interaction, and a significant positive effect of Condemnation in the No Information condition. As can be seen in the third and fourth columns of Table S1, these two results are qualitatively robust across all four measures. Interestingly, the effects appear stronger for the more "specific" measures (probability of not transgressing and specific trust) than for the more "general" measures (general trust and liking). See Conclusion section for a discussion of possible implications of these differences.

Table S1. Results by measure in Study 1

	Info X Condemn ANOVA			**Pairwise t-tests: effect of target condemning**	
Measure	**Info**	**Condemn**	**Info X Condemn**	**No info**	**Good info**
Prob. of not transgressing	F(1,615)=77.11, p<.001	F(1,615)=35.43, p<.001	F(1,615)=9.88, p=.002	t(301)=6.39, p<.001	t(314)=2.00, p=.046
Specific trust	F(1,615)=144.01, p<.001	F(1,615)=12.16, p<.001	F(1,615)=6.41, p=.012	t(301)=4.43, p<.001	t(314)=0.65, p=.514
General trust	F(1,615)=104.73, p<.001	F(1,615)=1.74, p=.188	F(1,615)=6.24, p=.013	t(301)=2.79, p=.006	t(314)=-0.81, p=.417
Liking	F(1,615)=89.60, p<.001	F(1,615)=2.02, p=.156	F(1,615)=3.19, p=.075	t(301)=2.42, p=.016	t(314)=-0.24, p=.807

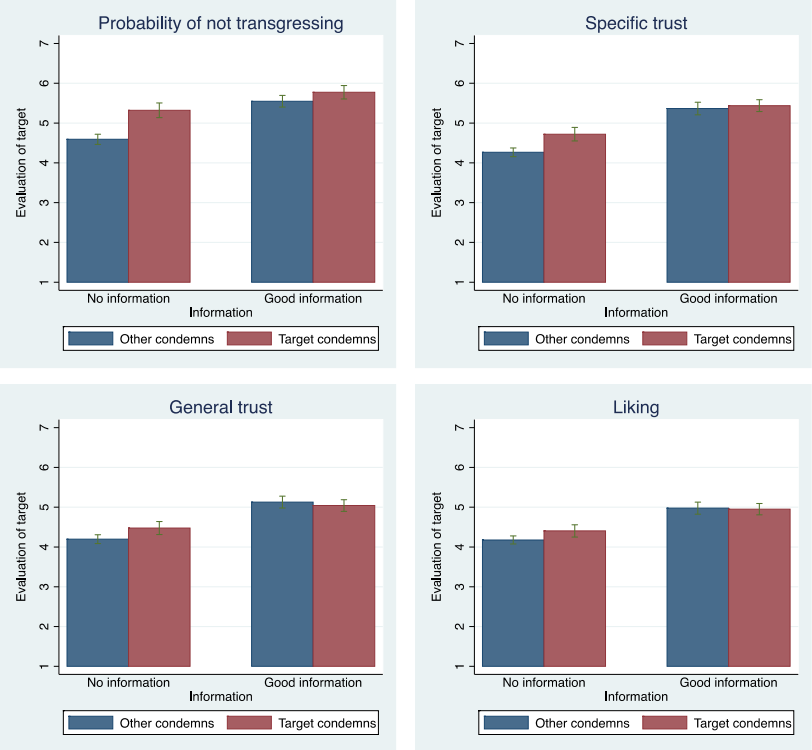


Figure S1. Results by measure in Study 1. Mean evaluation of targets by condition, shown separately for each dependent measure. Error bars represent 95% confidence intervals.

DOI: 10.1177/0956797616685771

Similarly, in Study 2, the key results, reported in the main text, were a significant positive Signal type X Signaling interaction, and a significant positive effect of Condemnation in the Target Signals condition. As can be seen in the third and fourth columns of Table S2, these two results are qualitatively robust across measures. As in Study 1, these effects appear stronger for the more "specific" measures (probability of not transgressing and specific trust) than the more "general" measures (general trust and liking). See Conclusion section for a discussion of possible implications of these differences.

Table S2. Results by measure in Study 2

	Signal type X Signaling ANOVA			**Pairwise t-tests: effect of condemnation**	
Measure	**Signal type**	**Signaling**	**Signal type X Signaling**	**Target signals**	**Other signals**
Prob. of not transgressing	F(1,799)=16.15, p<.001	F(1,799)=165.01, p<.001	F(1,799)=17.31, p<.001	t(397)=5.04, p<.001	t(402)=-0.12, p=.905
Specific trust	F(1,799)=8.21, p=.004	F(1,799)=154.83, p<.001	F(1,799)=11.44, p<.001	t(397)=3.95, p<.001	t(402)=-0.42, p=.674
General trust	F(1,799)=3.05, p=.081	F(1,799)=100.20, p<.001	F(1,799)=7.50, p=.006	t(397)=2.84, p=.005	t(402)=-0.81, p=.421
Liking	F(1,799)=2.86, p=.091	F(1,799)=133.24, p<.001	F(1,799)=4.10, p=.043	t(397)=2.27, p=.024	t(402)=-0.29, p=.774

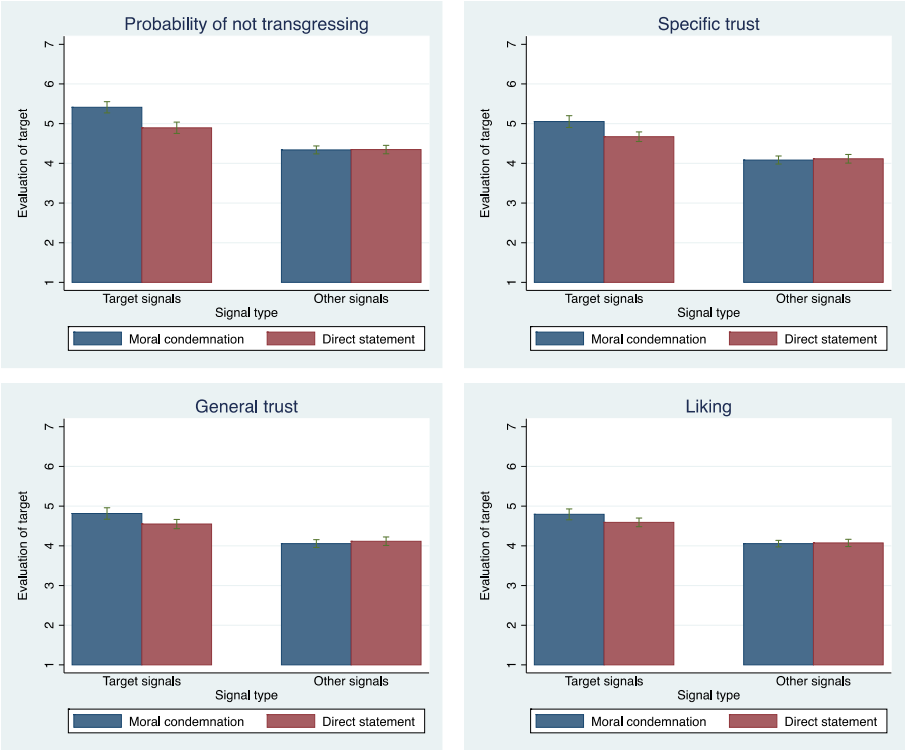


Figure S2. Results by measure in Study 2. Mean evaluation of targets by condition, shown separately for each dependent measure. Error bars represent 95% confidence intervals.

DOI: 10.1177/0956797616685771

The key Study 3 result, reported in the main text, was that hypocrites were judged more negatively than liars. As can be seen in the second column of Table S3, this result is directionally true for all measures, but is stronger for ratings of the target as a good person, and liking of the target, as compared to ratings of the target's trustworthiness and honesty. See Conclusion section for discussion of possible implications of these differences.

Table S3. Results by measure in Study 3

	Condition ANOVA	**Pairwise t-tests**	
Measure		**Hypocrite vs liar**	**Liar vs control transgressor**
Good person	F(2,448)=12.19, p<.001	t(299)=-2.57, p=.011	t(301)=-2.38, p=.018
Like	F(2,448)=25.28, p<.001	t(299)=-3.38, p<.001	t(301)=-3.77, p<.001
Trust	F(2,448)=27.66, p<.001	t(299)=-1.72, p=.086	t(301)=-5.38, p<.001
Honest	F(2,448)=33.44, p<.001	t(299)=-0.88, p=.378	t(301)=-6.83. p<.001

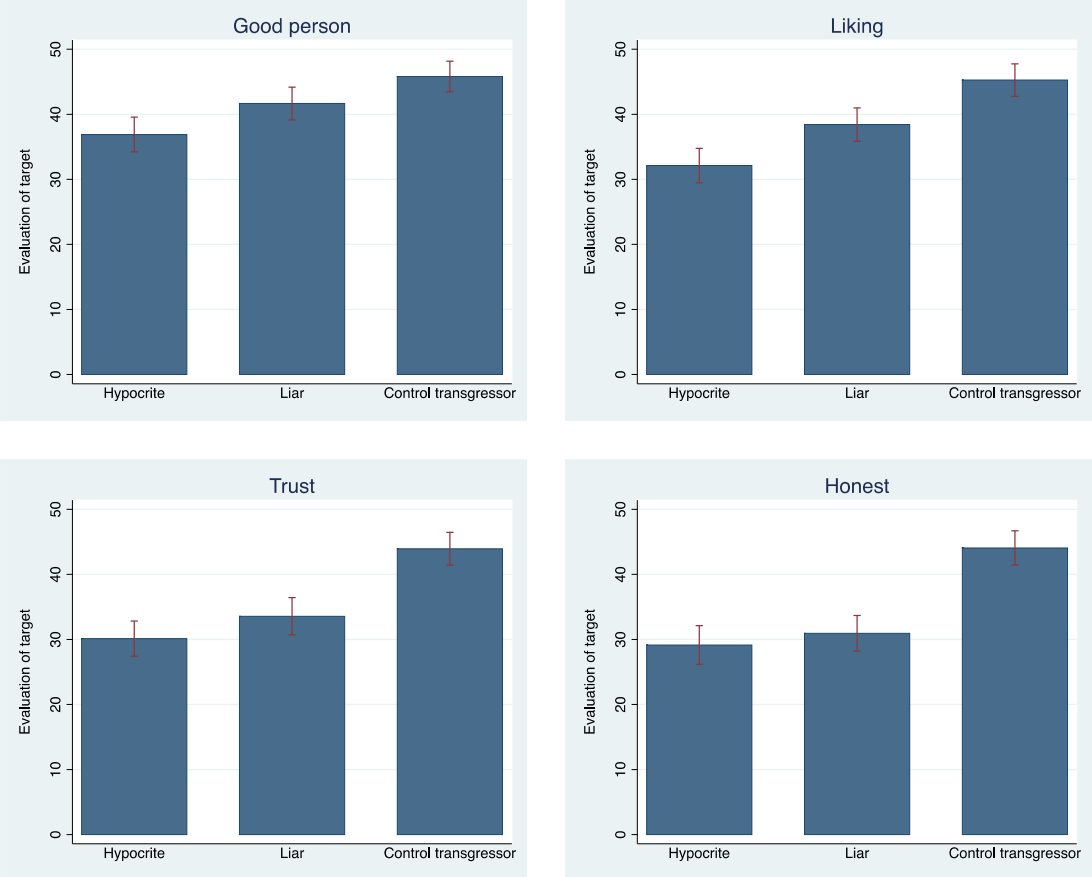


Figure S3. Results by measure in Study 3. Mean evaluation of targets by condition, shown separately for each dependent measure. Error bars represent 95% confidence intervals.

DOI: 10.1177/0956797616685771

The key Study 4 results, reported in the main text, were that traditional hypocrites were judged more negatively than honest hypocrites, and that honest hypocrites were judged no more negatively than control transgressors. As can be seen in the second and forth columns of Table S4, these results are robust across all measures (and when it comes to judgments of honesty, honest hypocrites are actually judged *more favorably* than control transgressors).

Table S4. Results by measure in Study 4

	Condition ANOVA	**Pairwise t-tests**		
Measure		**Traditional hypocrite vs honest hypocrite**	**Traditional hypocrite vs control transgressor**	**Honest hypocrite vs control transgressor**
Good person	F(2,449)=16.51, p<.001	t(301)=-5.32, p<.001	t(300)=-4.53, p<.001	t(297)=0.56, p=.579
Like	F(2,449)=17.47, p<.001	t(301)=-4.88, p<.001	t(300)=-5.20, p<.001	t(297)=-0.46, p=.647
Trust	F(2,449)=37.15, p<.001	t(301)=-7.60, p<.001	t(300)=-7.31, p<.001	t(297)=0.34, p=.731
Honest	F(2,449)=54.59, p<.001	t(301)=-9.57, p<.001	t(300)=-8.07, p<.001	t(297)=2.50, p=.013

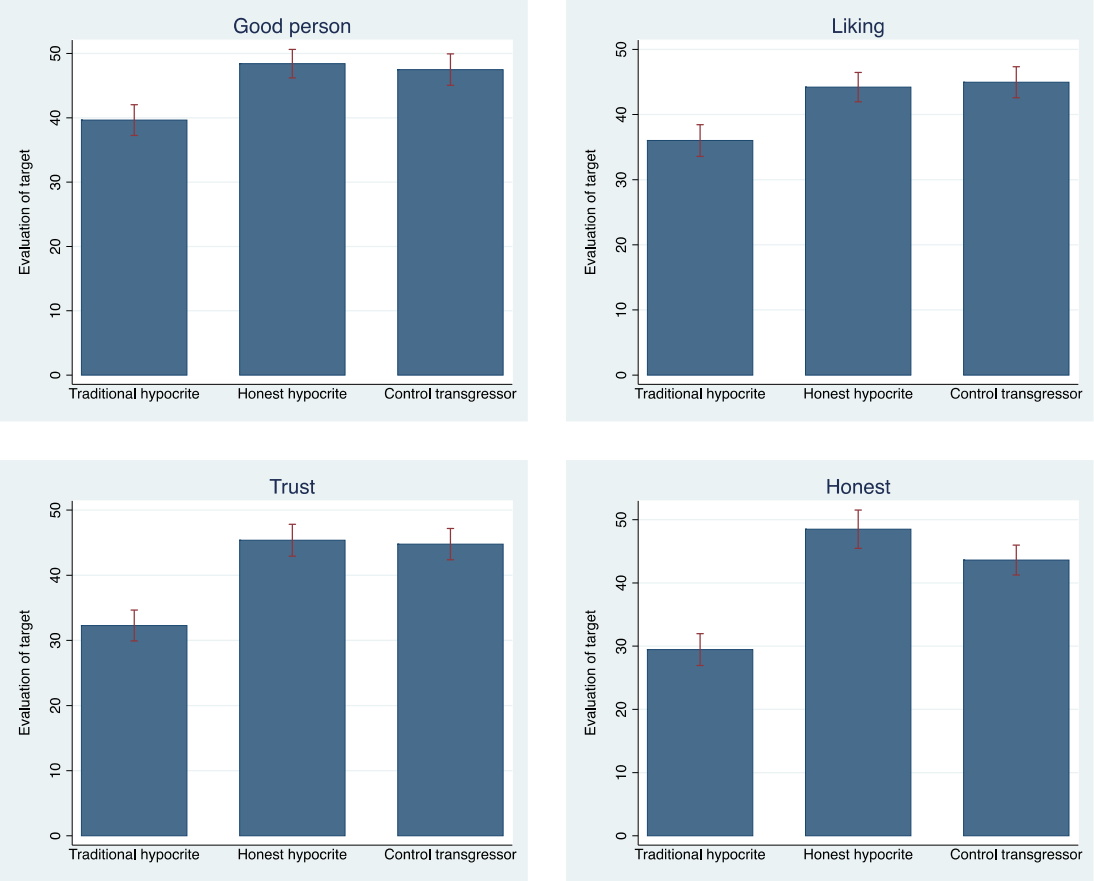


Figure S4. Results by measure in Study 4. Mean evaluation of targets by condition, shown separately for each dependent measure. Error bars represent 95% confidence intervals.

DOI: 10.1177/0956797616685771

The key Study 5 results, reported in the main text, were that disclosure hypocrites were judged more negatively than honest hypocrites, and that disclosure hypocrites were judged no more positively than traditional hypocrites. As can be seen in the fifth and sixth columns of Table S5, these results are robust across all measures.

Table S5. Results by measure in Study 5

	Condition ANOVA	**Pairwise t-tests**			
Measure		**Traditional hypocrite vs honest hypocrite**	**Honest hypocrite vs control transgressor**	**Disclosure hypocrite vs honest hypocrite**	**Disclosure hypocrite vs traditional hypocrite**
Good person	F(3,608)=18.67, p<.001	t(305)=-7.02, p<.001	t(303)=2.87, p=.004	t(304)=-6.00, p<.001	t(305)=0.95, p=.341
Like	F(3,608)=19.95, p<.001	t(305)=-6.42, p<.001	t(303)=0.98, p=.330	t(304)=-5.76, p<.001	t(305)=0.70, p=.484
Trust	F(3,608)=25.43, p<.001	t(305)=-6.73, p<.001	t(303)=0.43, p=.666	t(304)=-6.17, p<.001	t(305)=0.34, p=.733
Honest	F(3,608)=32.12, p<.001	t(305)=-8.84, p<.001	t(303)=2.20, p=.029	t(304)=-6.54, p<.001	t(305)=1.74, p=.082

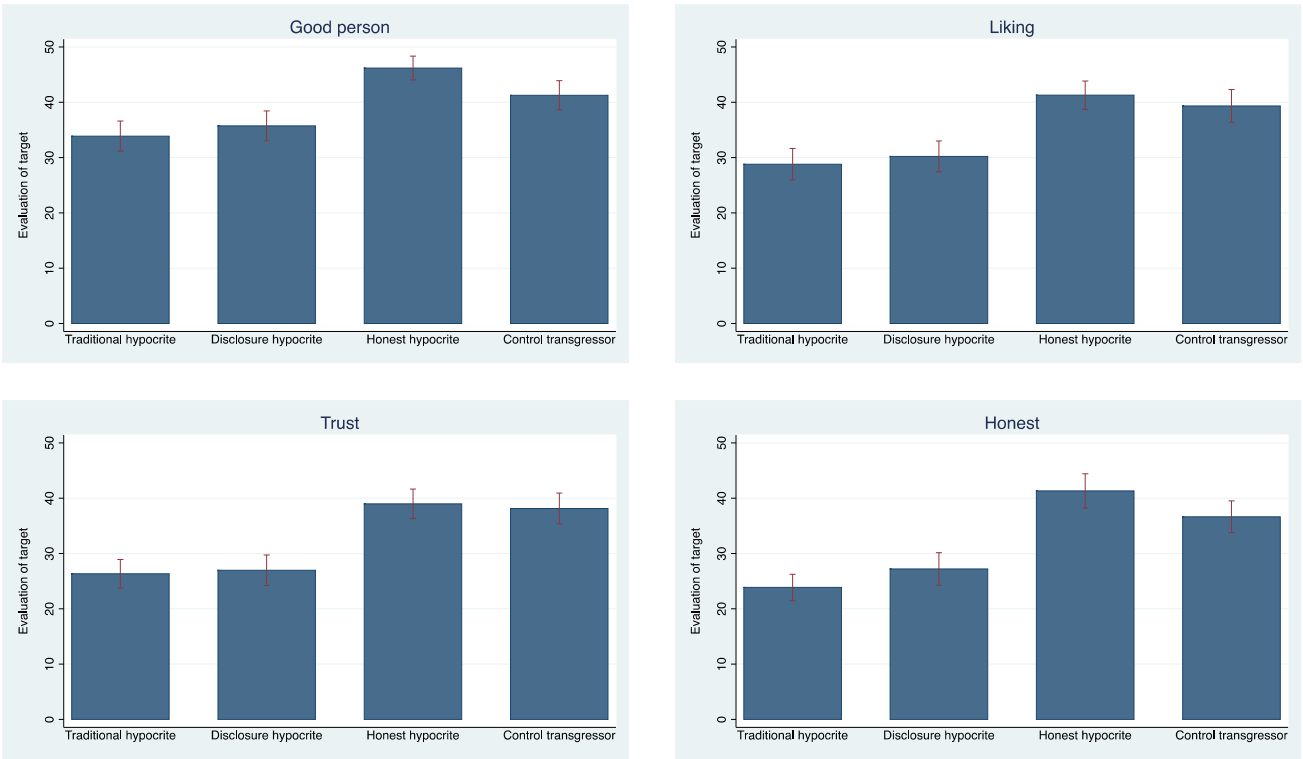


Figure S5. Results by measure in Study 5. Mean evaluation of targets by condition, shown separately for each dependent measure. Error bars represent 95% confidence intervals.

DOI: 10.1177/0956797616685771

Conclusions

The above analyses show that in each study, effects of condition differ significantly by dependent measure. However, they also show that the overall pattern of results is qualitatively similar across dependent measures. Here, we report on two theoretically interesting patterns of differences shown above, and discuss possible interpretations of these differences. However, because our analyses of differences between dependent measures were exploratory, further research should be conducted to replicate and interpret the patterns that we found.

First, in Study 1, the effect of condemnation in the "no information condition", and consequently the interaction between condemnation and information, is larger for the two more "specific" dependent measures (probability of condemnation and specific trust) than for the two more "general" dependent measures (general trust and liking). Similarly, in Study 2, the effect of condemnation in the "target signals" condition, and consequently the interaction between signaling and signal type, is larger for the two more "specific" dependent measures than for the two more "general" dependent measures. These patterns may be consistent with a mechanism by which condemnation serves as a somewhat domain specific signal of the target's behavior, which in secondarily (and thus more weekly) influences more general evaluations of the target.

Second, in Study 3, the contrast between hypocrites and liars is larger for ratings of the target as a good person, and liking of the target, as compared to ratings of the target's trustworthiness and honesty. As discussed in the discussion of the main text, this pattern is consistent with the tentative hypothesis that hypocrites may be judged as worse than liars for reasons beyond being more misleading—for example, because their false signaling may be more likely to create negative outcomes, such as maligning the condemned, shaming others into changing their behavior, and unfairly boosting the hypocrite's reputation.

DOI: 10.1177/0956797616685771

Further demographic analyses

In the main text, we report statistics on the gender and age composition of our samples for each study. Here, we report further demographic-related analyses. First, we plot descriptive statistics on all collected demographic variables for all studies. Second, we investigate possible interactions between demographic variables and condition effects for all studies.

Descriptive statistics

In our studies, we collected the following demographic variables: age, gender, income, and education. Below, we plot descriptive statistics for all of these demographic variables for each study. We note income and education were measured using categorical scales, with the options illustrated in the x-axis of the relevant plots. We also note that when plotting age distributions, for the purpose of visualizing our data, we dropped four total participants (across all five studies) who entered an age of under 10 years.

DOI: 10.1177/0956797616685771

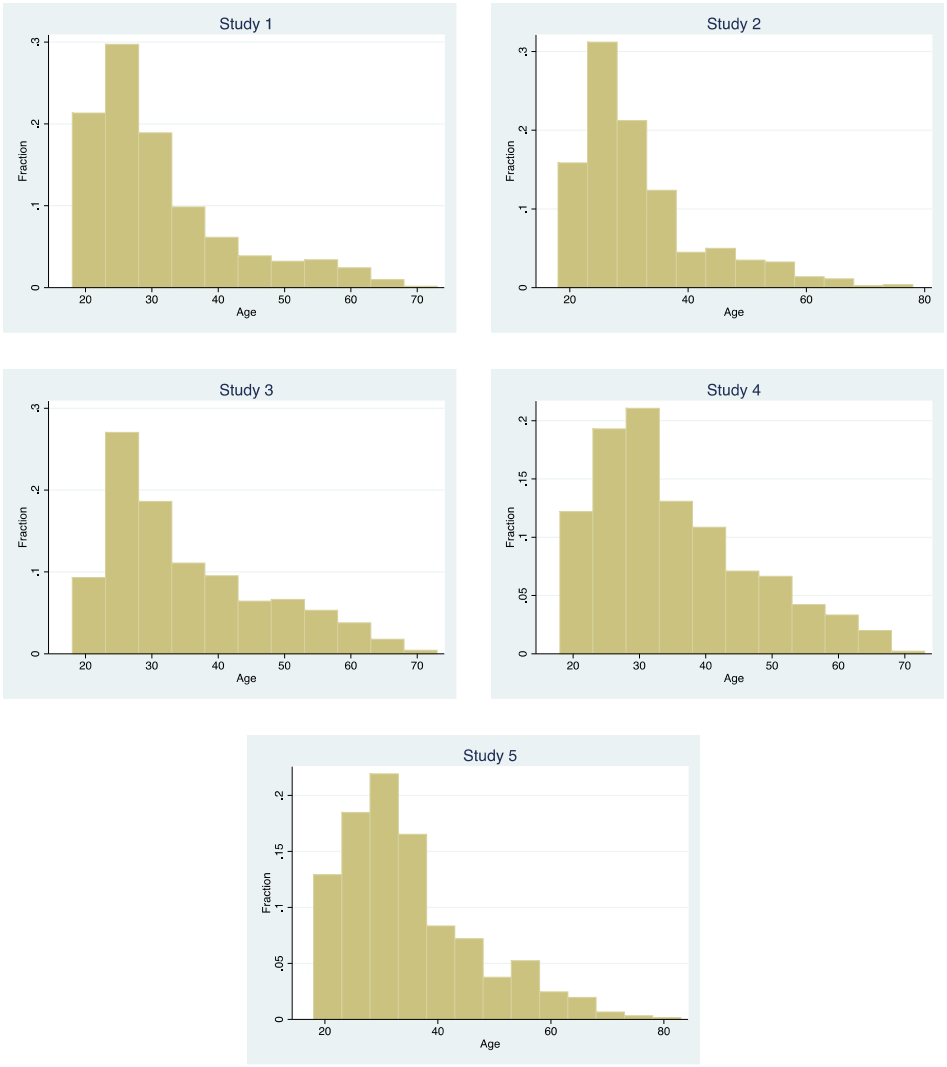


Figure S6. Subject age distribution across studies. Histograms show fraction of the subject pool in each age group for Studies 1-5.

DOI: 10.1177/0956797616685771

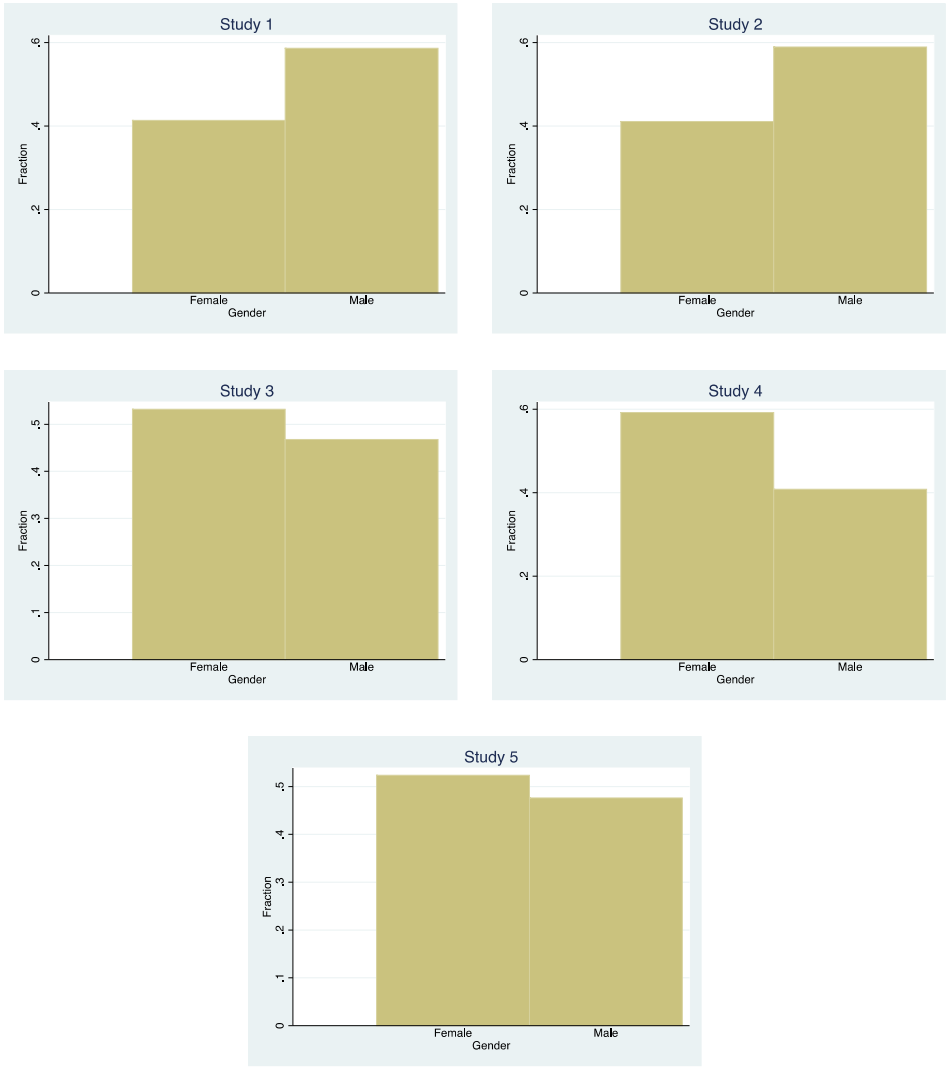


Figure S7. Subject gender distribution across studies. Histograms show fraction of the subject pool of each gender for Studies 1-5.

DOI: 10.1177/0956797616685771

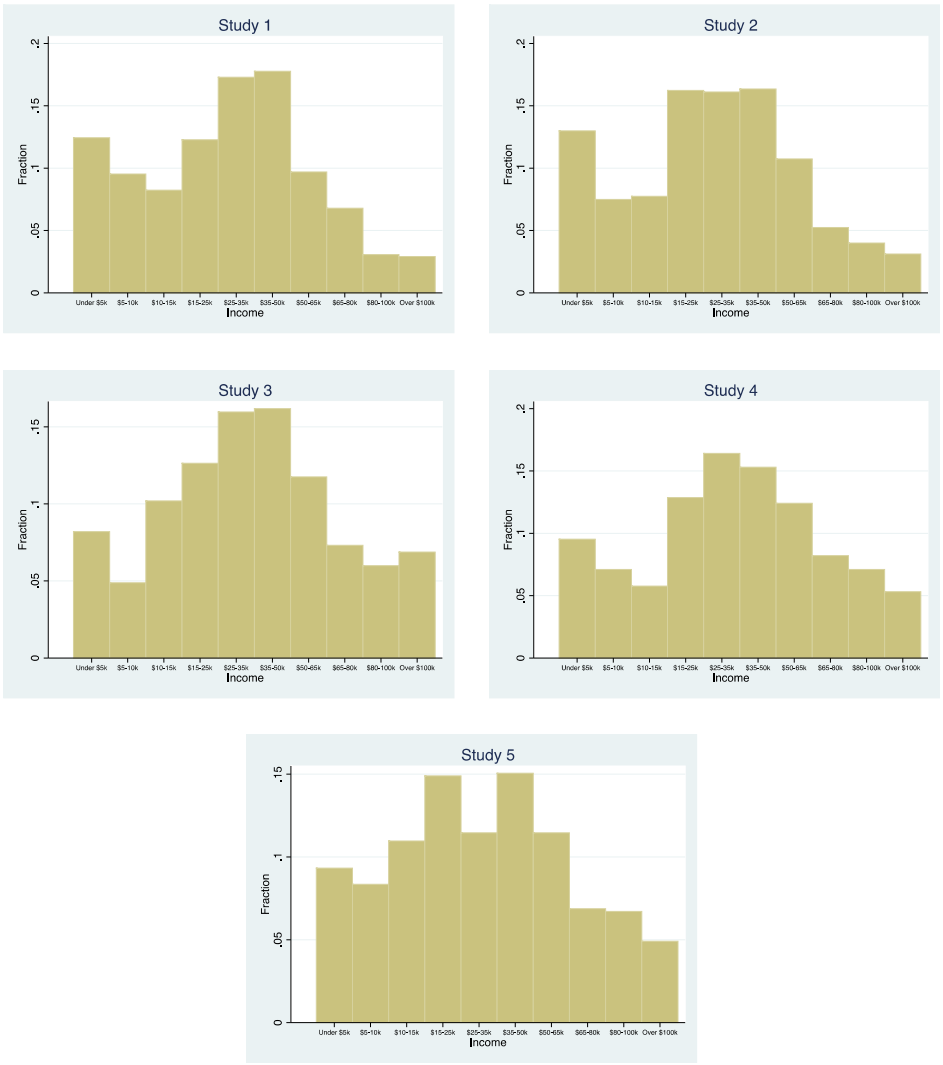


Figure S8. Subject income distribution across studies. Histograms show fraction of the subject pool in each income group for Studies 1-5.

DOI: 10.1177/0956797616685771

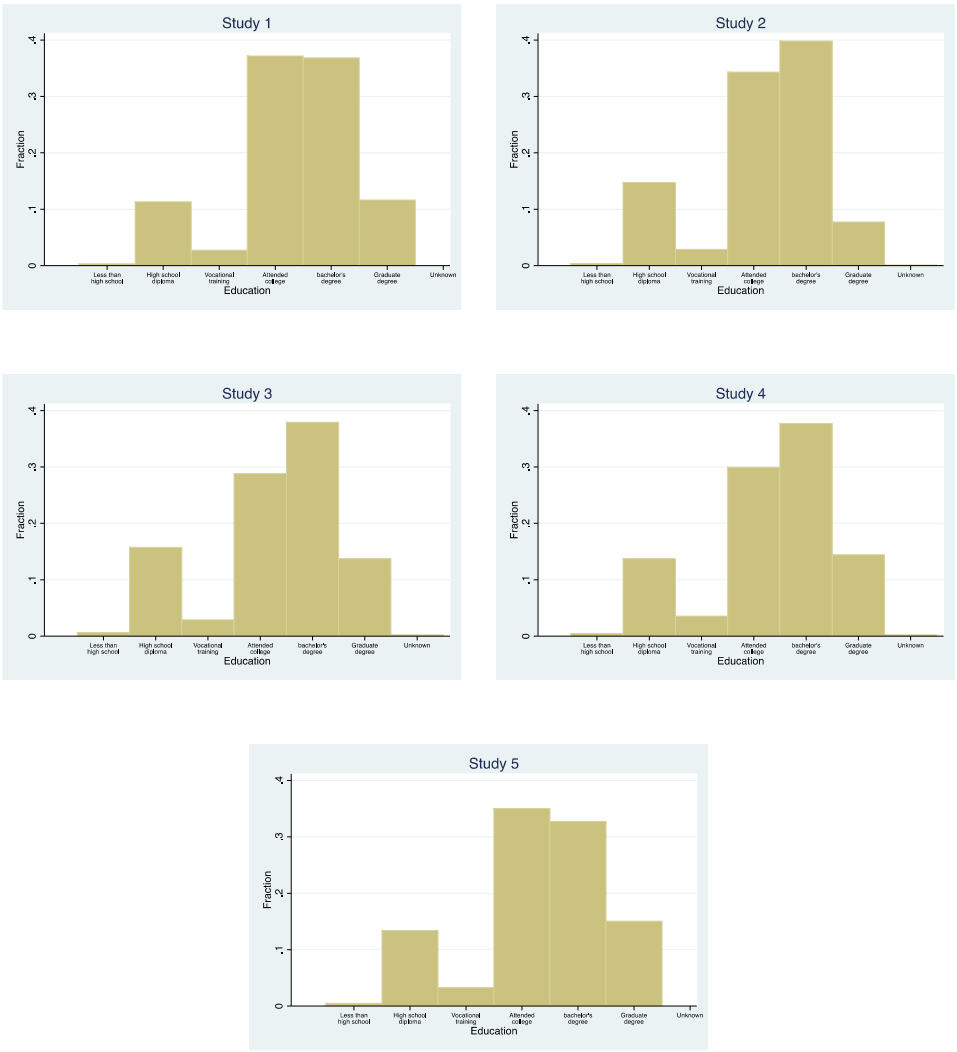


Figure S9. Subject education distribution across studies. Histograms show fraction of the subject pool in each education group for Studies 1-5.

Interaction effects

Next, we investigated potential interactions between condition variables and demographic variables in each study. When analyzing the data, we set our threshold for statistical significance at the standard cutoff of $p < .05$, despite the fact that we were conducting exploratory analyses with multiple comparisons. With this weak threshold for significance, we found only one (barely) statistically significant demographic interaction for one condition variable in one study (see below).

More specifically, in Studies 1 and 2, we analyzed gender and education as categorical variables. For each variable, we conducted ANOVAs with the two condition variables and the relevant demographic variable as factors, and investigated all main effects and interactions. We found no significant two- or three-way interactions for either gender or education. Next, we analyzed age and income as continuous variables. We conducted linear regressions with the two condition variables and the relevant demographic variable as independent variables, and investigated all main effects and interactions. The one significant effect we found was a negative two-way interaction between age and signaling, b = -0.01, SE = 0.01, p = .043. This effect indicates that within the moral condemnation condition, the effect of the target signaling is relatively larger for younger participants. Because this effect was not predicted and is only significant when the p-value is not corrected for the many exploratory tests we ran, we interpret this result with caution.

For Studies 3-5, we again analyzed gender and education as categorical variables. For each variable, we conducted ANOVAs with the condition variable and the relevant demographic variable as factors, and investigated both main effects and their interaction. We found no significant interaction for either variable. Next, we analyzed age and income as continuous variables. For each possible pairwise contrast between conditions, we conducted a linear regression with both condition and the relevant demographic variable as independent variables, and investigated both main effects and their interaction. We found no significant interaction for either variable for any condition contrast.

Together, these results suggest that within our Amazon Turk sample, our results are relatively robust across the demographic variables we measured. Further research should investigate whether this finding holds true across further demographic variables of interest, such

DOI: 10.1177/0956797616685771

as religiosity or ethnicity. Furthermore and perhaps more importantly, future research should investigate other populations outside of Amazon Turk, including non-"WEIRD" cultures (Henrich, Heine, & Norenzayan, 2010).

Full experimental stimuli

Study 1

In Study 1, subjects evaluate the following four gender-matched scenarios in a random order:

A) Drugs Scenario (female names: Brianna, Samantha, and Maria):

Imagine that you are an athlete on a track team. Recently, your coach has become concerned that members of the team are using an illegal performance-enhancing drug called Vitronil. Vitronil use threatens your team's eligibility to compete, and gives individual athletes unfair advantages.

[No information conditions: Two of your teammates are named **Brian** and **Sam**. You know nothing about if **Brain** uses Vitronil. You also know nothing about if **Sam** uses Vitronil.]
VS
[Good information conditions: Two of your teammates are named **Brian** and **Sam**. You overheard another member of the track team saying that **Brian** did not use Vitronil at his last track competition. In contrast, you know nothing about if **Sam** uses Vitronil.]

[Condemnation conditions: One day, you are having a conversation with **Brian**. You tell them a story about a mutual acquaintance, **Mark**, who is a competitive swimmer. After you finish your story, **Brian** mentions that he heard that **Mark** got caught using Vitronil right before an important swim meet. In telling his story, **Brian** expresses strong disapproval of Vitronil use.
VS
[No condemnation conditions: One day, you are having a conversation with **Sam**. You tell them a story about a mutual acquaintance, **Mark**, who is a competitive swimmer. After you finish your story, **Sam** mentions that he heard that **Mark** got caught using Vitronil right before an important swim meet. In telling his story, **Sam** expresses strong disapproval of Vitronil use.]

Comprehension questions (always presented in this fixed order):

DOI: 10.1177/0956797616685771

Who is **Brian**?

- ◯ A member of your team, and you know nothing about if he uses Vitronil
- ◯ A member of your team, and you overheard that he did not use Vitronil at his last track competition
- ◯ An acquaintance who used Vitronil

Who is **Sam**?

- ◯ A member of your team, and you know nothing about if he uses Vitronil
- ◯ A member of your team, and you overheard that he did not use Vitronil at his last track competition
- ◯ An acquaintance who used Vitronil

Who is **Mark**?

- ◯ A member of your team, and you know nothing about if he uses Vitronil
- ◯ A member of your team, and you overheard that he did not use Vitronil at his last track competition
- ◯ An acquaintance who used Vitronil

Who told you that **Mark** used Vitronil?

- ◯ Brian
- ◯ Sam
- ◯ Mark

<u>Dependent variables (presented in a random order):</u>

How much do you like **Brian**?

	1- Very little	2	3	4	5	6	7- Very much
	◯	◯	◯	◯	◯	◯	◯

How much would you generally trust **Brian** across contexts?

	1- Very little	2	3	4	5	6	7- Very much
	◯	◯	◯	◯	◯	◯	◯

How likely do you think **Brian** is to <u>use</u> Vitronil in the future?

	1- Very unlikely	2	3	4	5	6	7- Very likely
	◯	◯	◯	◯	◯	◯	◯

How much would you trust **Brian** as a competitor on your team?

	1- Very little	2	3	4	5	6	7- Very much
	◯	◯	◯	◯	◯	◯	◯

B) Work scenario (female names: Tina, Rachel, and Danielle):

DOI: 10.1177/0956797616685771

Imagine that you are an employee at an organization. At your organization, you have to work closely with partners on important projects. Individuals are evaluated on the basis of the joint work they produce with their partners.

[No information conditions: Two of your co-workers are named **Thomas** and **Rick**. You know nothing about what **Thomas** is like to work with as a partner. You also know nothing about what **Rick** is like to work with as a partner.]
VS
[Good information conditions: Two of your co-workers are named **Thomas** and **Rick**. You overheard another member of the organization saying that **Thomas** was a reliable partner in his last project at work. In contrast, you know nothing about what **Rick** is like to work with as a partner.]

[Condemnation conditions: One day, you are having a conversation with **Thomas**. You tell them a story about a mutual acquaintance, **Daniel**, who is an employee at a similar organization and also works with a partner. After you finish your story, **Thomas** mentions that he heard that **Daniel** failed to meet a critical deadline, causing him and his partner to perform poorly at an important presentation. In telling his story, **Thomas** expresses strong disapproval of unreliable partners.]
VS
[No condemnation conditions: One day, you are having a conversation with **Rick**. You tell them a story about a mutual acquaintance, **Daniel**, who is an employee at a similar organization and also works with a partner. After you finish your story, **Rick** mentions that he heard that **Daniel** failed to meet a critical deadline, causing him and his partner to perform poorly at an important presentation. In telling his story, **Rick** expresses strong disapproval of unreliable partners.]

[Subjects then answer comprehension questions and rate Thomas, using measures as in scenario (A)]

DOI: 10.1177/0956797616685771

C) Academic scenario (female names: Katie, Becca, and Gabrielle):

Imagine that you are a student in a chemistry course. The chemistry course involves difficult take-home exams. The exams are taken at home, but it is against the rules to use the Internet or discuss the exam with other members of the course. Recently, your instructor has become concerned that students are cheating on the exams.

[No information conditions: Two other members of the course are named **Kyle** and **Ben**. You know nothing about if **Kyle** cheats on his exams. You also know nothing about if **Ben** cheats on his exams.]
VS
[Good information conditions: Two other members of the course are named **Kyle** and **Ben**. You overheard another member of the course saying that **Kyle** did not cheat on his last exam. In contrast, you know nothing about if **Ben** cheats on his exams.]

[Condemnation conditions: One day, you are having a conversation with **Kyle.** You tell them a story about a mutual acquaintance, **Gabriel**, who is a law student. After you finish your story, **Kyle** mentions that he heard that **Gabriel** has been cheating in his law courses. In telling his story, **Kyle** expresses strong disapproval of academic cheating.]
VS
[No condemnation conditions: One day, you are having a conversation with **Ben.** You tell them a story about a mutual acquaintance, **Gabriel**, who is a law student. After you finish your story, **Ben** mentions that he heard that **Gabriel** has been cheating in his law courses. In telling his story, **Ben** expresses strong disapproval of academic cheating.]

[Subjects then answer comprehension questions and rate Kyle, using measures as in scenario (A)]

D) Romantic scenario (female names: Sarah, Jenny and Anna):

Imagine that you are a member of a hiking club. The hiking club is a great way to meet new people, including romantic partners, as new members join regularly and get to know each other on hikes.

[No information conditions: Two other members of the club are named **Steven** and **Josh**. You know nothing about what **Steven** is like as a romantic partner. You also know nothing about what **Josh** is like as a romantic partner.]
VS
[Good information conditions: Two other members of the club are named **Steven** and **Josh**. You overheard another member of the club saying that **Steven** was faithful in his last relationship. In contrast, you know nothing about what **Josh** is like as a romantic partner.]

[Condemnation conditions: One day, you are having a conversation with **Steven.** You tell them a story about a mutual acquaintance, **Adam**, who is in a serious relationship. After you finish your story, **Steven** mentions that he heard that **Adam** has been regularly cheating in his relationship. In telling his story, **Steven** expresses strong disapproval of romantic cheating.]
VS
[No condemnation conditions: One day, you are having a conversation with **Josh.** You tell them a story about a mutual acquaintance, **Adam**, who is in a serious relationship. After you finish your story, **Josh** mentions that he heard that **Adam** has been regularly cheating in his relationship. In telling his story, **Josh** expresses strong disapproval of romantic cheating.]

[Subjects then answer comprehension questions and rate Steven, using measures as in scenario (A)]

DOI: 10.1177/0956797616685771

Study 2

In Study 2, subjects evaluate the following four gender-matched scenarios in a random order:

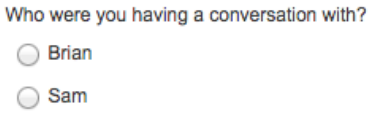
A) Drugs Scenario (female names: Brianna and Samantha):

Imagine that you are an athlete on a track team. Recently, your coach has become concerned that members of the team are using an illegal performance-enhancing drug called Vitronil. Vitronil use threatens your team's eligibility to compete, and gives individual athletes unfair advantages.

Two of your teammates are named **Brian** and **Sam**. You know nothing about if either of them use Vitronil.

[Signaling / condemnation condition: One day, you are having a conversation with **Brian**. The two of you are discussing how different members of your team compete at meets. Specifically, you are talking about who stays clean, and who takes Vitronil. In your discussion, **Brian** emphasizes that he disapproves of taking Vitronil.]
VS
[Control / condemnation condition: One day, you are having a conversation with **Sam**. The two of you are discussing how different members of your team compete at meets. Specifically, you are talking about who stays clean, and who takes Vitronil. In your discussion, **Sam** emphasizes that he disapproves of taking Vitronil.]
VS
[Signaling / direct statements condition: One day, you are having a conversation with **Brian**. The two of you are discussing how different members of your team compete at meets. Specifically, you are talking about who stays clean, and who takes Vitronil. In your discussion, **Brian** emphasizes that he does not take Vitronil.]
VS
[Control / direct statements condition: One day, you are having a conversation with **Sam**. The two of you are discussing how different members of your team compete at meets. Specifically, you are talking about who stays clean, and who takes Vitronil. In your discussion, **Sam** emphasizes that he does not take Vitronil.]

Comprehension question:



[Subjects then rate Brian, using the same measures as in Study 1.]

B) Work scenario (female names: Tina and Rachel):

Imagine that you are an employee at an organization. At your organization, you have to work closely with partners on important projects. Individuals are evaluated on the basis of the joint work they produce with their partners.

Two of your co-workers are named **Thomas** and **Rick**. You know nothing about what either of them are like to work with as partners.

[Signaling / condemnation condition: One day, you are having a conversation with **Thomas**. The two of you are discussing how different co-workers perform as partners. Specifically, you are talking about who is reliable and meets deadlines, and who is unreliable and fails to meet deadlines. In your discussion, **Thomas** emphasizes that he disapproves of unreliable partners.]
VS
[Control / condemnation condition: One day, you are having a conversation with **Rick**. The two of you are discussing how different co-workers perform as partners. Specifically, you are talking about who is reliable and meets deadlines, and who is unreliable and fails to meet deadlines. In your discussion, **Rick** emphasizes that he disapproves of unreliable partners.]
VS

DOI: 10.1177/0956797616685771

[Signaling / direct statements condition: One day, you are having a conversation with **Thomas**. The two of you are discussing how different co-workers perform as partners. Specifically, you are talking about who is reliable and meets deadlines, and who is unreliable and fails to meet deadlines. In your discussion, **Thomas** emphasizes that he is not an unreliable partner.]
VS
[Control / direct statements condition: One day, you are having a conversation with **Rick**. The two of you are discussing how different co-workers perform as partners. Specifically, you are talking about who is reliable and meets deadlines, and who is unreliable and fails to meet deadlines. In your discussion, **Rick** emphasizes that he is not an unreliable partner.]

[Subjects then answer comprehension question and rate Thomas, using measures as in scenario (A)]

C) Academic scenario (female names: Katie and Becca):
Imagine that you are a student in a chemistry course. The chemistry course involves difficult take-home exams. The exams are taken at home, but it is against the rules to use the Internet or discuss the exam with other members of the course. Recently, your instructor has become concerned that students are cheating on the exams.

Two other members of the course are named **Kyle** and **Ben**. You know nothing about if either of them cheat on their exams.

[Signaling / condemnation condition: One day, you are having a conversation with **Kyle.** The two of you are discussing how different students in your course complete their exams. Specifically, you are talking about who is honest and works alone, and who cheats by collaborating with others and using the Internet. In your discussion, **Kyle** emphasizes that he disapproves of cheating on exams.]
VS
[Control / condemnation condition: One day, you are having a conversation with **Ben.** The two of you are discussing how different students in your course complete their exams. Specifically, you are talking about who is honest and works alone, and who cheats by collaborating with others and using the Internet. In your discussion, **Ben** emphasizes that he disapproves of cheating on exams.]
VS
[Signaling / direct statements condition: One day, you are having a conversation with **Kyle.** The two of you are discussing how different students in your course complete their exams. Specifically, you are talking about who is honest and works alone, and who cheats by collaborating with others and using the Internet. In your discussion, **Kyle** emphasizes that he does not cheat on his exams.]
VS
[Control / direct statements condition: One day, you are having a conversation with **Ben.** The two of you are discussing how different students in your course complete their exams. Specifically, you are talking about who is honest and works alone, and who cheats by collaborating with others and using the Internet. In your discussion, **Ben** emphasizes that he does not cheat on his exams.]

[Subjects then answer comprehension question and rate Kyle, using measures as in scenario (A)]

D) Romantic scenario (female names: Sarah and Jenny):

Imagine that you are a member of a hiking club. The hiking club is a great way to meet new people, including romantic partners, as new members join regularly and get to know each other on hikes.

Two other members of the club are named **Steven** and **Josh**. You know nothing about what either of them are as in romantic relationships.

[Signaling / condemnation condition: One day, you are having a conversation with **Steven.** The two of you are discussing what different members of your club are as in romantic relationships. Specifically, you are talking about

DOI: 10.1177/0956797616685771

who is trustworthy and stays faithful, and who is not trustworthy and cheats on their partners. In your discussion, **Steven** emphasizes that he disapproves of cheating in relationships.]
VS
[Control / condemnation condition: One day, you are having a conversation with **Josh.** The two of you are discussing what different members of your club are as in romantic relationships. Specifically, you are talking about who is trustworthy and stays faithful, and who is not trustworthy and cheats on their partners. In your discussion, **Josh** emphasizes that he disapproves of cheating in relationships.]
VS
[Signaling / direct statements condition: One day, you are having a conversation with **Steven.** The two of you are discussing what different members of your club are as in romantic relationships. Specifically, you are talking about who is trustworthy and stays faithful, and who is not trustworthy and cheats on their partners. In your discussion, **Steven** emphasizes that he does not cheat in relationships.]
VS
[Control / direct statements condition: One day, you are having a conversation with **Josh.** The two of you are discussing what different members of your club are as in romantic relationships. Specifically, you are talking about who is trustworthy and stays faithful, and who is not trustworthy and cheats on their partners. In your discussion, **Josh** emphasizes that he does not cheat in relationships.]

[Subjects then answer comprehension question and rate Steven, using measures as in scenario (A)]

DOI: 10.1177/0956797616685771

Study 3

In Study 3, subjects evaluate the following four scenarios in a random order:

A) Music Scenario

Becky and her friend Amanda are discussing a mutual acquaintance. Amanda mentions that the acquaintance often downloads music illegally from the Internet.

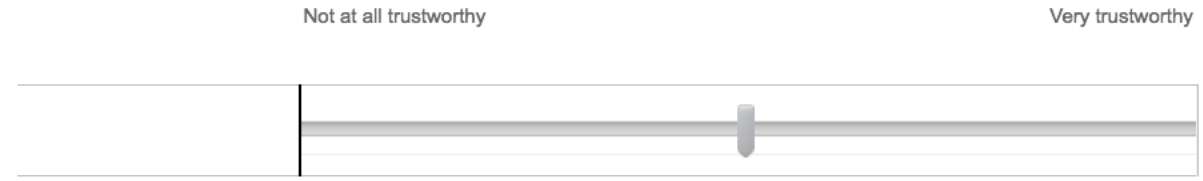
[Hypocrisy condition: Becky says that she thinks it is morally wrong to download music illegally from the Internet.]
VS
[Liar condition: Becky says that she doesn't download music illegally from the Internet.]
VS
[No sentence presented in the control transgressor condition.]

Shortly after their conversation, Becky goes online, and downloads music illegally.

DOI: 10.1177/0956797616685771

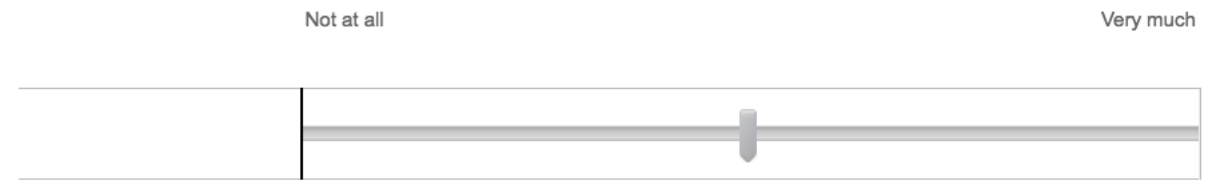
Dependent variables (presented in random order):

How **trustworthy** do you think Becky is?



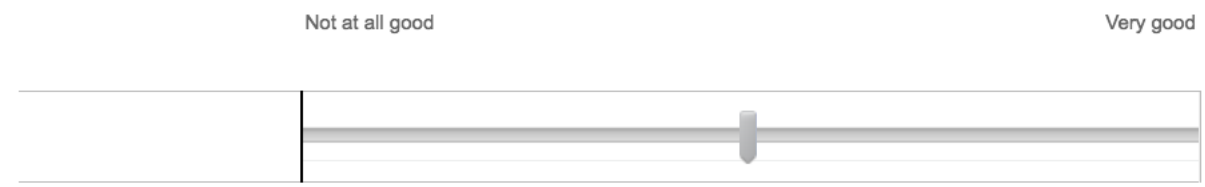
How **hypocritical** do you think Becky is?

How much do you **like** Becky?



How **honest** do you think Becky is?

How **good a person** do you think Becky is?



DOI: 10.1177/0956797616685771

B) Jury Scenario

Jennifer and her friend Rose are discussing a mutual acquaintance. Rose mentions that the acquaintance recently tried to get out of jury duty.

[Hypocrisy condition: Jennifer says that she thinks it is morally wrong to try to get out of jury duty.]
VS
[Liar condition: Jennifer says that she doesn't try to get out of jury duty.]
VS
[No sentence presented in the control transgressor condition.]

Shortly after their conversation, Jennifer gets called for jury duty, and tries to get out of it.

[Subjects then rate Jennifer, using measures as in scenario (A)]

C) Phone Call Scenario

Bruce and his friend are Zach are discussing a mutual acquaintance. Zach mentions that the acquaintance often ignores his mother's phone calls.

[Hypocrisy condition: Bruce says that he thinks it is morally wrong to ignore your mother's phone calls.]
VS
[Liar condition: Bruce says that he doesn't ignore his mother's phone calls.]
VS
[No sentence presented in the control transgressor condition.]

Shortly after their conversation, Bruce notices that his mother is calling, and ignores the call.

[Subjects then rate Bruce, using measures as in scenario (A)]

D) Printing Scenario

Kevin and his friend Jack are discussing a mutual acquaintance. Jack mentions that the acquaintance often uses a lot of paper by printing documents single-sided.

[Hypocrisy condition: Kevin says that he thinks it is morally wrong to use a lot of paper by printing documents single-sided.]
VS
[Liar condition: Kevin says that he doesn’t use a lot of paper by printing documents single-sided.]
VS
[No sentence presented in the control transgressor condition.]

Shortly after their conversation, Kevin has a large document to print, and uses a lot of paper by printing it single-sided.

[Subjects then rate Kevin, using measures as in scenario (A)]

DOI: 10.1177/0956797616685771

Study 4

Study 4 is identical to Study 3, except that the "liar" condition is replaced with an "honest hypocrisy" condition (described below), and for clarity we consequently rename the "hypocrisy" condition the "traditional hypocrisy" condition. The Music Scenario of Study 4 looks as follows:

Becky and her friend Amanda are discussing a mutual acquaintance. Amanda mentions that the acquaintance often downloads music illegally from the Internet.

[Traditional hypocrisy condition: Becky says that she thinks it is morally wrong to download music illegally from the Internet.]
VS
[Honest hypocrisy condition: Becky says that she thinks it is morally wrong to download music illegally from the Internet, but that she sometimes does it anyway.]
VS
[No sentence presented in the control transgressor condition.]

Shortly after their conversation, Becky goes online, and downloads music illegally.

Dependent variables are also the same as in Study 3; and other scenarios are also the same as in Study 3, with analogous changes as in the Music Scenario shown above.

DOI: 10.1177/0956797616685771

Study 5

In Study 5, subjects evaluate the "Music and Phone Call" and "Jury and Printing" scenarios in random order, and then evaluate the targets using the same dependent variables as in Studies 3-4. These scenarios involve counterbalancing (which we collapse over in our analyses), as follows:

When presented with the "Music and Phone Call" scenario, we randomly assign subjects to either see the "Condemn Music" or "Condemn Phone Call" version. Furthermore, and orthogonally, we also randomly assign them to see the "Music First" or "Phone Call First" version.

Likewise, when presented with the "Jury and Printing" scenario, we randomly assign subjects to either see the "Condemn Jury" or "Condemn Printing" version. Furthermore, and orthogonally, we also randomly assign them to see the "Jury First" or "Printing First" version.

A) Music and Phone Call Scenario

Condemn Music, Music First Version

Becky and her friend Amanda are discussing issues in their lives, like downloading music and answering their parents' phone calls.

[Traditional hypocrisy condition: Becky tells Amanda that she thinks it is morally wrong when people download music illegally from the Internet.]
VS
[Disclosure hypocrisy condition: Becky tells Amanda that she thinks it is morally wrong when people download music illegally from the Internet, but that she sometimes ignores her mother's phone calls.]
VS
[Honest hypocrisy condition: Becky tells Amanda that she thinks it is morally wrong when people download music illegally from the Internet, but that she sometimes does it anyway.]
VS
[No sentence presented in the control transgressor condition.]

Shortly after their conversation, Becky goes online, and downloads music illegally. She also notices that her mother is calling, and ignores the call.

Condemn Music, Phone Call First Version

Becky and her friend Amanda are discussing issues in their lives, like answering their parents' phone calls and downloading music.

[Traditional hypocrisy condition: Becky tells Amanda that she thinks it is morally wrong when people download music illegally from the Internet.]
VS
[Disclosure hypocrisy condition: Becky tells Amanda that she thinks it is morally wrong when people download music illegally from the Internet, but that she sometimes ignores her mother's phone calls.]
VS
[Honest hypocrisy condition: Becky tells Amanda that she thinks it is morally wrong when people download music illegally from the Internet, but that she sometimes does it anyway.]
VS
[No sentence presented in the control transgressor condition.]

Shortly after their conversation, Becky notices that her mother is calling, and ignores the call. She also goes online, and downloads music illegally.

DOI: 10.1177/0956797616685771

Condemn Phone Call, Music First Version

Becky and her friend Amanda are discussing issues in their lives, like downloading music and answering their parents' phone calls.

[Traditional hypocrisy condition: Becky tells Amanda that she thinks it is morally wrong when people ignore their mothers' phone calls.]
VS
[Disclosure hypocrisy condition: Becky tells Amanda that she thinks it is morally wrong when people ignore their mothers' phone calls, but that she sometimes downloads music illegally from the Internet.]
VS
[Honest hypocrisy condition: Becky tells Amanda that she thinks it is morally wrong when people ignore their mothers' phone calls, but that she sometimes does it anyway.]
VS
[No sentence presented in the control transgressor condition.]

Shortly after their conversation, Becky goes online, and downloads music illegally. She also notices that her mother is calling, and ignores the call.

Condemn Phone Call, Phone Call First Version

Becky and her friend Amanda are discussing issues in their lives, like answering their parents' phone calls and downloading music.

[Traditional hypocrisy condition: Becky tells Amanda that she thinks it is morally wrong when people ignore their mothers' phone calls.]
VS
[Disclosure hypocrisy condition: Becky tells Amanda that she thinks it is morally wrong when people ignore their mothers' phone calls, but that she sometimes downloads music illegally from the Internet.]
VS
[Honest hypocrisy condition: Becky tells Amanda that she thinks it is morally wrong when people ignore their mothers' phone calls, but that she sometimes does it anyway.]
VS
[No sentence presented in the control transgressor condition.]

Shortly after their conversation, Becky notices that her mother is calling, and ignores the call. She also goes online, and downloads music illegally.

B) Jury and Printing Scenario

Condemn Jury, Jury First Version

Kevin and his friend Jack are discussing issues in their lives, like attending jury duty and printing documents.

[Traditional hypocrisy condition: Kevin tells Jack that he thinks it is morally wrong when people try to get out of jury duty.]
VS
[Disclosure hypocrisy condition: Kevin tells Jack that he thinks it is morally wrong when people try to get out of jury duty, but that he sometimes uses a lot of paper by printing documents single-sided.]
VS
[Honest hypocrisy condition: Kevin tells Jack that he thinks it is morally wrong when people try to get out of jury duty, but that he sometimes does it anyway.]
VS
[No sentence presented in the control transgressor condition.]

Shortly after their conversation, Kevin gets called for jury duty, and tries to get out of it. He also has a large document to print, and uses a lot of paper by printing it single-sided.

DOI: 10.1177/0956797616685771

Condemn Jury, Printing First Version

Kevin and his friend Jack are discussing issues in their lives, like printing documents and attending jury duty.

[Traditional hypocrisy condition: Kevin tells Jack that he thinks it is morally wrong when people try to get out of jury duty.]
VS
[Disclosure hypocrisy condition: Kevin tells Jack that he thinks it is morally wrong when people try to get out of jury duty, but that he sometimes uses a lot of paper by printing documents single-sided.]
VS
[Honest hypocrisy condition: Kevin tells Jack that he thinks it is morally wrong when people try to get out of jury duty, but that he sometimes does it anyway.]
VS
[No sentence presented in the control transgressor condition.]

Shortly after their conversation, Kevin has a large document to print, and uses a lot of paper by printing it single-sided. He also gets called for jury duty, and tries to get out of it.

Condemn Printing, Jury First Version

Kevin and his friend Jack are discussing issues in their lives, like attending jury duty and printing documents.

[Traditional hypocrisy condition: Kevin tells Jack that he thinks it is morally wrong when people use a lot of paper by printing documents single-sided.]
VS
[Disclosure hypocrisy condition: Kevin tells Jack that he thinks it is morally wrong when people use a lot of paper by printing documents single-sided, but that he sometimes tries to get out of jury duty.]
VS
[Honest hypocrisy condition: Kevin tells Jack that he thinks it is morally wrong when people use a lot of paper by printing documents single-sided, but that he sometimes does it anyway.]
VS
[No sentence presented in the control transgressor condition.]

Shortly after their conversation, Kevin gets called for jury duty, and tries to get out of it. He also has a large document to print, and uses a lot of paper by printing it single-sided.

Condemn Printing, Printing First Version

Kevin and his friend Jack are discussing issues in their lives, like printing documents and attending jury duty.

[Traditional hypocrisy condition: Kevin tells Jack that he thinks it is morally wrong when people use a lot of paper by printing documents single-sided.]
VS
[Disclosure hypocrisy condition: Kevin tells Jack that he thinks it is morally wrong when people use a lot of paper by printing documents single-sided, but that he sometimes tries to get out of jury duty.]
VS
[Honest hypocrisy condition: Kevin tells Jack that he thinks it is morally wrong when people use a lot of paper by printing documents single-sided, but that he sometimes does it anyway.]
VS
[No sentence presented in the control transgressor condition.]

Shortly after their conversation, Kevin has a large document to print, and uses a lot of paper by printing it single-sided. He also gets called for jury duty, and tries to get out of it.

DOI: 10.1177/0956797616685771

References

Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world? *Behavioral and brain sciences, 33*(2-3), 61-83.