“I’m Just Being Honest.” When and Why Honesty Enables Help Versus Harm

Emma Levine and David Munguia Gomez
University of Chicago Booth School of Business

Although honesty is typically conceptualized as a virtue, it often conflicts with other equally important moral values, such as avoiding interpersonal harm. In the present research, we explore when and why honesty enables helpful versus harmful behavior. Across 5 incentive-compatible experiments in the context of advice-giving and economic games, we document four central results. First, honesty enables selfish harm: people are more likely to engage in and justify selfish behavior when selfishness is associated with honesty than when it is not. Second, people are selectively honest; people are more likely to be honest when honesty is associated with selfishness than when honesty is associated with altruism. Third, these effects are more consistent with genuine, rather than motivated, preferences for honesty. Fourth, even when individuals have no selfish incentive to be honest, honesty can lead to interpersonal harm because people avoid information about how their honest behavior affects others. This research unearths new insights on the mechanisms underlying moral choice, and consequently, the contexts in which moral principles are a force of good versus a force of evil.

Keywords: honesty, ethics, deception, justification, information avoidance

Supplemental materials: http://dx.doi.org/10.1037/pspi0000242.supp

Honesty is typically seen as a virtue. This notion is deeply engrained in society and is reinforced across myriad religious and philosophical texts, as well as corporate codes of conduct and practical advice. Dell, for example, begins their code of conduct with the proclamation, “We are honest” (Dell Code of Conduct, 2012), and articles in the New York Times feature advice such as, “The secret to success in business and in life is to never, ever, ever tell a lie” (Campbell, 2014). Psychological research on interpersonal trust and ethics also reinforces this advice, suggesting that honesty is essential for personal and relational fulfillment (Fowers, 2005; Griffin, 1967).

Individuals’ positive judgments of honesty stem in part from the association between honesty and prosociality. People typically conflate honesty and benevolence (Levine & Schweitzer, 2014), and conceptualize honest acts as both truthful and promoting social welfare. Honesty, however, is not always benevolent. Consider giving feedback to a colleague who you believe has poor presentation skills and with whom you are competing for a promotion. Although honestly providing this feedback may be motivated by prosocial intentions such as a desire to help your colleague improve, it may also be motivated by less noble intentions such as a desire to undermine their confidence. If you believed that your honest feedback would cause public embarrassment and make you look smarter, would you be using honesty for good? Or rather, is it possible that you might be motivated to cause harm, and feel it is appropriate to do so, because you can justify your behavior by simply saying, “I was just being honest?” In the present research, we shed light on these questions by exploring how having the opportunity to tell the truth influences both selfishness and altruism. In doing so, we provide novel theoretical and practical insights into how honesty, and moral principles broadly, enable harmful and helpful behaviors.

Theory

Honesty as an Enabler of Selfishness

Individuals are often motivated to behave selfishly (Green & Cowden, 1992; Schwartz, 1986) and harm others (Bushman & Huesmann, 2010). However, acting on these motivations conflicts with individuals’ fundamental desire to think of themselves as good and ethical people (Greenwald, 1980; Mazar, Amir, & Ariely, 2008; Santitioso, Kunda, & Fong, 1990). Thus, to behave selfishly or antisocially and maintain a positive self-image, individuals often seek to attribute their behavior to moral or, at the very least, amoral motives (e.g., Batson, Kobrynowicz, Dinner-
Selective Honesty

Because moral principles allow people to justify selfish acts, people tend to selectively follow moral principles. That is, people are more likely to endorse a given principle when the principle aligns with their self-interest than when it does not. For example, political liberals are more likely to endorse consequentialism in a trolley dilemma when consequentialism aligns with their political views (e.g., when consequentialism promotes the lives of Black Americans rather than White Americans; Uhlmann, Pizarro, Tannenbaum, & Ditto, 2009). Similarly, people are more likely to be honest when it helps them than when it harms them (Erat & Gneezy, 2012). Consistent with these findings, we hypothesize that:

Hypothesis 2: People are more likely to be honest when honesty is selfish than when honesty is altruistic.

Mechanisms Underlying the Honesty-Selfishness Relationship

Although honesty could certainly operate as a justification for selfishness, consistent with the logic outlined above, this is not the only psychological process that could link honesty with increased selfishness. Consider a person who is more likely to be honest when it helps them than when it harms them. However, it is also possible that this pattern of behavior characterizes someone who genuinely values both self-interest and honesty. That is, people can have a genuine preference for the principle and follow it to different degrees in different settings.

To test whether a moral principle is genuinely motivating behavior (vs. being used to justify a self-serving behavior), a control condition that does not involve that principle is needed. Consider again a person who is more likely to be honest when it helps them than when it harms them. If that person has a preference for self-interest over altruism (when honesty is not relevant), their tendency to be honest more often when honesty is selfish than altruistic may simply reflect this underlying preference. Adding honesty as a potential motivation for behavior may increase selfish and altruistic behavior at equal rates. Such a pattern of behavior would reveal a genuine (i.e., stable) desire to adhere to a principle of honesty, even if the rate of honesty differs across conditions. Alternatively, if adding honesty as a potential motivation for behavior increased the rate of selfish behavior more than it increased the rate of altruistic behavior, this would suggest that honesty served as a justification for selfishness.

We visualize these comparisons in Figure 1. Existing research on self-serving justifications typically compares how people behave when a moral principle could or could not motivate their self-serving behavior. For example, scholars have examined people’s propensity to engage in self-serving deception when it does or does not benefit a third party (Gino, Ayal, & Ariely, 2013; Wiltermuth, 2011). This is the equivalent of Effect 1 in Figure 1 (Hypothesis 1).

Figure 1. A (hypothetical) illustration of how to test for self-serving versus genuine preferences for honesty. Test typically performed in past research on how moral principles function as self-serving justifications. "Test typically performed in past research on how people selectively use moral principles. " In the present research, we compare Effect 1 to Effect 3 to test the extent to which moral principles function as self-serving justifications versus genuine motivators of behavior.
A related body of research on the motivated use of moral principles (e.g., Ditto, Pizarro, & Tannenbaum, 2009; Uhlmann et al., 2009) has examined whether people are more likely to endorse a specific principle when the principle aligns with their self-interest than when it does not. This comparison is the equivalent of Effect 2 in Figure 1: it compares the likelihood of endorsing or following a moral principle when doing so aligns or conflicts with one’s own interests. In the context of honesty, we refer to this phenomenon as “selective honesty” (Hypothesis 2).

As discussed above, however, it is possible that one could follow a principle to different degrees based on whether the principle aligns with their self-interest, despite being genuinely motivated by that principle. To test whether a moral principle serves as a justification or a genuine motivator of behavior it is necessary to compare Effect 1 (the effect of the principle on self-serving behavior) to Effect 3 (the effect of the principle on nonself-serving behavior). In the context of honesty, if the effect of honesty on selfishness is greater than the effect of honesty on altruism, then this provides evidence that honesty serves as a justification of selfish behavior, rather than a genuine motivator of behavior.

Notably, only one article that we know of uses a design similar to what we propose to tease apart self-serving and genuine motives for following moral principles. In Study 3 of Gino et al. (2013), the authors compare individuals’ likelihood of cheating when it only helps the self to the likelihood of cheating when it helps the self and others (Effect 1, adding a moral justification to self-interest), as well as individuals’ likelihood of cheating when it only helps others (Effect 3, adding a moral justification, but removing self-interest). The authors find that people are not more likely to cheat to help (only) others than they are to cheat to help themselves, suggesting that the presence of a beneficiary primarily acts as a justification of selfish behavior, rather than a genuine motivator of behavior.

However, existing research provides mixed evidence for genuine versus self-serving preferences for honesty. Although influential work on hypocrisy suggests that “principles may be used more reactively than proactively, more to justify or condemn action than to motivate it” (Batson et al., 1997, p. 1336), more recent research suggests that people do have intrinsic preferences for honesty. For example, a significant proportion of people are unwilling to tell lies that help themselves and others, suggesting that people do have a genuine aversion toward lying (and/or a genuine preference for honesty; Erat & Gneezy, 2012; see also Lundquist, Ellingsen, Gribbe, & Johannesson, 2009; Vanberg, 2008).

Given the mixed nature of existing research on honesty, we set up competing hypotheses on whether genuine versus self-serving preferences underlie the effects of honesty on selfish harm. Specifically, we test the validity of the following three hypotheses:

**Hypothesis 3a:** People are genuinely motivated by honesty. Honesty enables selfishness and altruism to a similar degree.

**Hypothesis 3b:** People use honesty to justify selfishness. Honesty enables selfishness, but, not altruism.

**Hypothesis 3c:** People have both genuine and self-serving preferences for honesty. Honesty enables both selfishness and altruism, but enables selfishness to a greater degree.

We test these hypotheses by comparing people’s propensity to engage in selfish versus altruistic behavior, when each behavior is or is not associated with honesty (consistent with Figure 1).

### Honesty and Information Avoidance

In addition to testing the mechanisms that underlie the relationship between honesty and selfish harm, we propose and test a new mechanism that may link honesty with harm, even in the absence of self-interest: information avoidance. Specifically, we propose that honesty, and moral principles broadly, can enable harm by increasing actors’ propensity to avoid learning about the social consequences of their actions.

Individuals often avoid information that puts them in compromising ethical positions. For example, Dana and colleagues (2007) found that participants playing a modified dictator game often chose not to find out how their selfish choice affected others, in order to behave selfishly without feeling bad about it. In their studies, participants had to choose between two allocations in a dictator game: a self-interested allocation that gave them $6 and their partner $1 and a fair allocation that gave each person $5. Some participants knew their payoff and their partner’s payoff; these participants chose the fair allocation the majority of the time. Other participants knew only their own payoff but had the option to find out their partner’s payoff (without any cost or effort) before choosing an allocation. Almost half of these participants chose not to see how their allocation choice affected their partner. As a result, these individuals were more likely to make the self-interested choice (i.e., choose the largest possible allocation for themselves) than individuals who had full information about the allocations, and presumably felt happier and less conflicted about this choice (Berman & Small, 2012). People also avoid information that could tempt them to be selfish. For example, people may avoid finding out the personal costs of cooperation to ensure that they cooperate with others and to signal this commitment to others (Hoffman, Yoeli, & Nowak, 2015; Jordan, Hoffman, Nowak, & Rand, 2016).

Outside of the moral domain, individuals avoid information to protect their intuitive preferences. For example, participants avoid learning how much money a boring task will earn them, and how many calories are in a tempting dessert, to protect their intuitive preference to complete a fun task and indulge, respectively (Woolley & Risen, 2018). Participants do actually care about this information; the same individuals who avoided the information use it when it is provided. However, individuals avoid the relevant information both to protect their emotions and simplify the decision. Taken together, these studies suggest that individuals may avoid information whenever learning such information would make them feel conflicted about their choice.

Learning that one’s honest thoughts or behaviors would harm others is likely to create a similar decision conflict. If one were to find out that an honest action would cause harm to others, then they would be forced to choose between the lesser of two evils—to harm or to lie. Harm is naturally construed as immoral (Gray, 1967).
Schein, & Ward, 2014; Schein & Gray, 2016, 2018) and, therefore, honestly causing harm is likely to elicit guilt and feelings of immorality. However, deception is also seen as immoral and, therefore, lying to help others is likely to elicit similar feelings (Levine et al., 2018). Rather than choose to harm or to lie, individuals can circumvent this moral conflict entirely by choosing not to find out about the harm caused by one’s honesty. They can just be honest, allowing themselves to feel moral without confronting the social consequences of their behavior. Therefore, we propose that:

Hypothesis 4: People are more likely to avoid finding out how their behavior affects others when their behavior is honest.

Even if information avoidance is driven by a genuine preference to follow the principle of honesty, this behavior can be problematic because it may cause decision-makers to take courses of action that undermine social welfare.

Overview of Studies

We test our hypotheses across five incentive-compatible experiments. In Study 1, we test Hypotheses 1, 2, and 3. Specifically, we test whether honesty enables selfishness, whether people are selectively honest, and whether self-serving or genuine preferences for honesty underlie these effects. In Studies 1, 3, and 4, we use modified deception games (e.g., Gneezy, 2005; Levine & Schweitzer, 2014, 2015). These games allow us to cleanly manipulate whether choices are associated with honesty or not, while holding all other aspects of the choice constant. In Study 2, we replicate our tests of Hypotheses 1, 2, and 3 using a more naturalistic paradigm; we study honesty within the context of advice-giving. In Study 3, we disentangle selfishness from interpersonal harm. We test whether honesty enables selfish, harmful acts primarily by increasing the degree to which selfishness seems justified or by decreasing the moral cost of harming others. Finally, in Studies 4 and 5, we test whether honesty causes actors to avoid looking at the social consequences of their actions.

The Institutional Review Board (IRB) of the University of Pennsylvania (Emma Levine’s former institution) approved all aspects of Studies 1 and 4, and the IRB of the University of Chicago approved all aspects of Studies 2, 3, and 5. Across all of our studies, sample sizes were determined in advance, and no conditions, participants or variables were dropped from any analyses. Studies 2, 3, and 5 were preregistered (see online supplemental materials 1.4 for details). Materials, data, and syntax are available at https://osf.io/gz4xh/.

Contribution

These studies deepen our theoretical and practical understanding of both honesty and morality. Although existing research has demonstrated that people selectively use moral principles in self-serving ways (e.g., Batson et al., 1997; Ditto et al., 2009; Gino et al., 2013), existing research has not examined how people selectively use honesty in particular. Given that honesty is one of the most commonly encountered moral values in everyday moral decision-making (Hofmann, Wisneski, Brandt, & Skitka, 2014; Iyer, 2010), investigating the selective use of honesty is practically important. Understanding when and why people speak up, rather than stay silent or engage in deception, allows practitioners to design incentives and structure choice sets in ways that encourage honesty across a range of dilemmas, including whistle-blowing (Waytz, Dungan, & Young, 2013), the disclosure of conflicts of interest (Loewenstein, Sah, & Cain, 2012), and the communication of negative news (Rosen & Tesser, 1970).

Moral situations involving honesty also have unique psychological properties that help us understand the nature of moral choices, more broadly. In many situations that involve honesty, moral actors are not simply faced with a choice to be honest or dishonest (i.e., to be moral or immoral); they can also say nothing (or engage in omission). Omission can serve as a moral compromise option when honesty is costly. In the current article, we develop new insights on how moral compromise options allow people to adjudicate between the desire to advance their self-interest, avoid interpersonal harm, and maintain a positive self-image.

We contribute to existing research on the selective use of moral principles not only by studying this topic in the context of honesty, but also by deepening our basic understanding of why selective morality occurs. We introduce new methodologies for testing why selective morality occurs. Just as Mullen and Monin (2016) clarified the importance of including control conditions for studying the nature of moral licensing versus consistency, we clarify the importance of including control conditions for studying the nature of selective morality. We provide a clear set of empirical tests (see Figure 1) that allow researchers to examine whether a moral principle is serving as a justification versus a genuine motivator of behavior, and we conduct these tests in the context of honesty. As we will reveal in our studies, we find that, surprisingly, selective rule-following is more consistent with genuine, than motivated, preferences for moral rules, at least in the present context. Finally, we introduce new mechanisms linking honesty with social harm by documenting how honesty can lead people to avoid information about the social consequences of their choices.

Study 1: The Effects of Honesty on Selfishness Versus Altruism

In Study 1, participants made decisions about how to allocate money between themselves and a partner. Specifically, they chose to either maintain a fair initial allocation, switch to a more selfish allocation, or switch to a more altruistic allocation. We manipulated whether the selfish allocation, the altruistic allocation, or neither allocation was associated with honesty.

Study 1 tests Hypothesis 1, that people are more likely to engage in selfish acts when they can do so honestly, as well as Hypothesis 2, that people are more likely to be honest when honesty is selfish than when it is altruistic. Study 1 also compares the validity of three competing hypotheses that could explain these effects. If people have a genuine and stable preference for honesty, we would expect honesty to increase the frequency of selfish and altruistic behaviors at equal rates (Hypothesis 3a). Alternatively, if people simply use honesty to justify selfishness, we would expect individuals to engage in selfish behavior, but not altruistic behavior, more frequently when it is associated with honesty (Hypothesis 3b). However, if people have a genuine preference for honesty, but also use it to justify selfishness, we would expect honesty to increase the frequency of both selfish and altruistic behaviors, but
to have a greater effect on selfishness versus altruism (Hypothesis 3c).

**Method**

**Participants.** We aimed to recruit 300 adults to participate in an online study via Amazon Mechanical Turk (MTurk), based on the a priori goal of recruiting 100 participants per condition. 336 participants started the study, but 23 were automatically kicked out of the study (before the manipulation) for failing an attention check, and an additional eight left the survey before completing our main measures. We ended up with a final sample of 305 adults (33% female; \( M_{\text{age}} = 30 \text{ years}, SD = 8.61 \)) who were eligible for analysis. In Study 1, we removed all incomplete observations from the dataset before any analyses.

**Procedure and materials.** In Study 1, we randomly assigned participants to one of three conditions in a between-subjects design: **Control**, **Selfish Honesty**, or **Altruistic Honesty**. Participants played an economic game, in which they were assigned to the role of Decider and paired with a Receiver (a future study participant). In the Decider, participants had the opportunity to allocate money between themselves and the Receiver.

All participants started the game with an initial allocation: \( $0.25 \) for the Decider (the participant) and \( $0.25 \) for the Receiver. In Study 1, we labeled the initial allocation “the default allocation” and told participants they could choose to keep the default or switch the allocation to one of two possible options. Participants had the opportunity to change the allocation to one of the following options:

- **Option A:** \( $0.55 \) for the Decider and \( $0.05 \) for the Receiver (selfish)
- **Option B:** \( $0.05 \) for the Decider and \( 0.55 \) for the Receiver (altruistic)

In the Control condition, participants had three choices: they could do nothing and maintain the initial allocation (which we conceptualize as inaction), they could switch the allocation to Option A, or they could switch the allocation to Option B. In the Selfish Honesty and Altruistic Honesty conditions, we told participants that the new allocation depended on their choice and the outcome of a random number generator that would generate a number between 1 and 9. Participants in these conditions learned the actual number generated, and then had to report whether the number was ODD or EVEN. If participants reported that the number was ODD, the allocation would switch to Option A (selfish); if they reported that the number was EVEN, the allocation would switch to Option B (altruistic). Participants could also choose not to report a number, which would maintain the initial allocation. Keeping the initial allocation reflects inaction via the omission of information; the participant could do and say nothing and receive the initial allocation. We label the decision to do nothing as “omission” rather than “inaction” in our Selfish and Altruistic Honesty conditions, because doing nothing necessarily required omitting information. Across our studies, “omission” refers to the omission of information, not an inaction that yields the same consequences as an equivalent action (as omission is typically defined in work on the omission bias; e.g., Baron & Ritov, 1994).

In the Selfish Honesty condition, the number was ODD. Therefore, participants in the Selfish Honesty condition had the opportunity to either (a) do nothing (omission) and receive the initial allocation, (b) honestly report that the number was ODD and receive Option A (selfish), or (c) dishonestly report that the number was EVEN and receive Option B (altruistic). This choice set models a situation in which an individual makes a discretionary choice to engage in hurtful (and self-serving) honesty. Participants could choose to say nothing and maintain the status quo or they could actively choose to voice honest information that harms others and helps themselves (tell a selfish truth). In our paradigm, participants could also choose to voice dishonest information that helps others and harms themselves (tell an altruistic lie). This design differs from related work (e.g., Batson et al., 1997) in that participants did not have a choice about whether to generate a random number in the first place, which would have allowed them to create the appearance of impartiality. Instead, participants simply learned the outcome of the random number generator, which we were able to verify for each participant, and had to choose whether or not to report it, or say nothing and maintain the status quo.

In the Altruistic Honesty condition, participants learned that the random number generator had generated an EVEN number. Therefore, participants in the Altruistic Honesty condition had the opportunity to either (a) do nothing (omission) and receive the initial allocation, (b) dishonestly report that the number was ODD and receive Option A (selfish lie), or (c) honestly report that the number was EVEN and receive Option B (Altruistic Honesty). That is, they had the choice to stay silent and maintain the status quo, tell a selfish lie, or tell an altruistic truth.

**Dependent variables.** The primary dependent variable was participants’ choice of the initial allocation (inaction/omission), Option A (selfish), or Option B (altruistic). Participants also rated their choice, motives, and identity:

**Ethical decision.** Participants rated how ethical and moral their choice was (\( r[305] = .95, p < .001 \)). We used 7-point Likert scales anchored at 1 = very unethical (immoral) and 7 = very ethical (moral).

**Justifications.** Participants also indicated the motivation for their choice. We measured participants’ honest justification using the following two items (\( r[305] = .82, p < .001 \)): “I made my choice because it was honest” and “I made my choice because I care about the truth.” We measured participants’ selfish justification using three items (\( \alpha = .74 \)): “I made my choice because it helped me,” “I made my choice because I wanted the best outcome for myself,” and “I made my choice because it helped others” (reverse-scored). All items were anchored at 1 = strongly disagree and 7 = strongly agree.

**Moral identity.** Participants then rated their moral identity using the following four items (\( \alpha = .95 \)): “I am an ethical person,” “I have good ethical judgment,” “Having good character is an important part of my sense of self,” and “Acting with integrity is part of who I am.” The latter two items were adapted from Aquino.
and Reed (2002). All items were anchored at 1 = strongly disagree and 7 = strongly agree.

**Preferred outcome.** At the end of the study, we asked participants to choose which of the outcomes (the initial allocation, Option A, or Option B) was most attractive to them. The purpose of this item was to understand whether our manipulation influenced the attractiveness of different outcomes, or if it simply enabled different behavior.

After participants submitted their responses, we collected demographic information and asked participants what they thought the purpose of the study was. Participants then received a bonus payment based upon their decisions.

**Analytical approach.** In each study, we test Hypothesis 1 and Hypothesis 3 using multinomial logistic regression, which allows us to test how Selfish Honesty and Altruistic Honesty each influenced participants’ choices among all three choice options (the altruistic allocation, inaction omission, and the selfish allocation). In our multinomial regressions, we ordered these choices from most costly to least costly such that choosing an altruistic allocation was coded as 1, choosing inaction (omission) was coded as 2, and choosing a selfish allocation was coded as 3. We treated inaction (omission) as our base outcome. Therefore, the regression test how Selfish Honesty and Altruistic Honesty each shifted people’s choices away from inaction (omission) and toward either the selfish or altruistic allocation. In our regressions, we dummy coded Selfish Honesty as 1 = Selfish Honesty condition, 0 = otherwise, and Altruistic Honesty as 1 = Altruistic Honesty condition, 0 = Otherwise.

We also conducted a simple chi-square test of proportions to test Hypothesis 2, whether people are more likely to be honest in the Selfish Honesty than the Altruistic Honesty condition.

### Results

**Preferred outcome.** The decision context did not affect the perceived attractiveness of each allocation (ps > .60). Table 1, Panel B and Figure 2, Panel B display these results.

**Choice.** However, the decision context significantly influenced participants’ allocation choices. Table 1, Panel A and Figure 2, Panel A depicts these results.

First, consistent with Hypothesis 1, we find that participants’ distribution of choices was significantly different in the Selfish Honesty condition and the Control condition ($\chi^2 = 22.42, p < .001$). Relative to the Control condition, participants in the Selfish

### Table 1

**Multinomial Logistic Regression on Choices in Study 1**

<table>
<thead>
<tr>
<th>Choice</th>
<th>Condition</th>
<th>$b$</th>
<th>SE</th>
<th>$p$</th>
<th>Lower</th>
<th>Upper</th>
</tr>
</thead>
<tbody>
<tr>
<td>Altruistic allocation</td>
<td>Altruistic Honesty</td>
<td>1.368</td>
<td>0.547</td>
<td>0.012</td>
<td>0.296</td>
<td>2.440</td>
</tr>
<tr>
<td></td>
<td>Selfish Honesty</td>
<td>-0.683</td>
<td>1.128</td>
<td>0.545</td>
<td>-2.894</td>
<td>1.529</td>
</tr>
<tr>
<td></td>
<td>Constant</td>
<td>-2.262</td>
<td>0.470</td>
<td>&lt;.001</td>
<td>-3.183</td>
<td>-1.341</td>
</tr>
<tr>
<td>Selfish allocation</td>
<td>Altruistic Honesty</td>
<td>-0.095</td>
<td>0.299</td>
<td>0.750</td>
<td>-0.681</td>
<td>0.491</td>
</tr>
<tr>
<td></td>
<td>Selfish Honesty</td>
<td>1.462</td>
<td>0.326</td>
<td>&lt;.001</td>
<td>0.823</td>
<td>2.102</td>
</tr>
<tr>
<td></td>
<td>Constant</td>
<td>&lt;.001</td>
<td>0.204</td>
<td>1.000</td>
<td>-0.400</td>
<td>0.400</td>
</tr>
</tbody>
</table>

**Post hoc tests**

**Overall effect of Altruistic Honesty**

$\chi^2 = 7.30, p = .026$

**Overall effect of Selfish Honesty**

$\chi^2 = 22.42, p < .001$

**Effect of Altruistic Honesty on Altruism vs. Effect of Selfish Honesty on Selfishness**

$\chi^2 = .02, p = .875$.

<table>
<thead>
<tr>
<th>Choice</th>
<th>Condition</th>
<th>$b$</th>
<th>SE</th>
<th>$p$</th>
<th>Lower</th>
<th>Upper</th>
</tr>
</thead>
<tbody>
<tr>
<td>Altruistic allocation</td>
<td>Altruistic Honesty</td>
<td>-13.292</td>
<td>755.428</td>
<td>0.986</td>
<td>-1493.903</td>
<td>1467.319</td>
</tr>
<tr>
<td></td>
<td>Selfish Honesty</td>
<td>0.656</td>
<td>1.255</td>
<td>0.601</td>
<td>-1.804</td>
<td>3.116</td>
</tr>
<tr>
<td></td>
<td>Constant</td>
<td>-3.258</td>
<td>1.019</td>
<td>0.001</td>
<td>-5.255</td>
<td>-1.261</td>
</tr>
<tr>
<td>Selfish allocation</td>
<td>Altruistic Honesty</td>
<td>-0.024</td>
<td>0.320</td>
<td>0.939</td>
<td>-0.651</td>
<td>0.603</td>
</tr>
<tr>
<td></td>
<td>Selfish Honesty</td>
<td>-0.051</td>
<td>0.320</td>
<td>0.873</td>
<td>-0.679</td>
<td>0.577</td>
</tr>
<tr>
<td></td>
<td>Constant</td>
<td>1.046</td>
<td>0.228</td>
<td>&lt;.001</td>
<td>0.599</td>
<td>1.493</td>
</tr>
</tbody>
</table>

**Post hoc tests**

**Overall effect of Altruistic Honesty**

$\chi^2 = .01, p = .997$

**Overall effect of Selfish Honesty**

$\chi^2 = .34, p = .843$

**Effect of Altruistic Honesty on Altruism vs. Effect of Selfish Honesty on Selfishness**

$\chi^2 < .001, p = .986$.

**Note.** CI = confidence interval. In this multinomial regression, we ordered choices from most costly to least costly such that choosing an altruistic allocation was coded as 1, choosing inaction (omission) was coded as 2, and choosing a selfish allocation was coded as 3. We treated inaction (omission) as our base outcome. Therefore, the regression tests how Selfish Honesty and Altruistic Honesty each shifted people’s choices away from inaction (omission) and toward either the selfish or altruistic allocation. We dummy coded Selfish Honesty as 1 = Selfish Honesty condition, 0 = otherwise, and Altruistic Honesty as 1 = Altruistic Honesty condition, 0 = Otherwise.
Honesty condition shifted their choices away from inaction and toward the selfish allocation, as revealed by the coefficients of the multinomial logistic regression ($b_{H11005} = 1.46$, $95\%$ confidence interval, CI $[0.82, 2.10]$; this effect represents Effect 1 in Figure 1).

We also find that participants’ distribution of choices was significantly different in the Altruistic Honesty condition and the Control condition ($b_{H9273} = 7.30$, $p < .026$). Relative to the Control condition, participants in the Altruistic Honesty condition shifted their choices away from inaction and toward the altruistic allocation ($b_{H11005} = 1.37$, $95\%$ CI $[0.30, 2.44]$; this effect represents Effect 3 in Figure 1).

Though having the opportunity to tell a selfish truth (Selfish Honesty condition) had a qualitatively larger effect on participants’ distribution of choices than having the opportunity to tell an altruistic truth (Altruistic Honesty condition) did, comparing the effect of the Selfish Honesty condition on selfish choice to the effect of Altruistic Honesty condition on altruistic choice (i.e., comparing Effect 1 to Effect 3 per Figure 1) reveals that effect sizes are not significantly different ($\chi^2 = 6.29$, $p = .448$). In other words, though people are more likely to be honest when it helps them than when it harms them (consistent with Hypothesis 2), we cannot reject the possibility that honesty is genuinely motivating behavior.

### Attitudinal measures.
We focus our attitudinal analyses only on participants who selected Option A (the selfish allocation), to test how linking selfishness with honesty influenced justifications. In Study 1, we found main effects of condition on Ethical Decision ($F(2, 167) = 67.84$, $p < .001$, $\eta_p^2 = .448$), Honest Justification ($F(2, 167) = 72.53$, $p < .001$, $\eta_p^2 = .465$), Selfish Justification ($F(2, 167) = 15.81$, $p < .001$, $\eta_p^2 = .159$), and Moral Identity ($F(2, 167) = 14.21$, $p < .001$, $\eta_p^2 = .145$). Individuals who made selfish choices judged their decision as more ethical, judged themselves as more motivated by honesty and less motivated by selfishness, and rated themselves higher in moral identity in the Selfish Honesty condition, relative to the Control and Altruistic Honesty conditions (all pairwise comparisons between Selfish Honesty condition and the Control and Altruistic Honesty conditions: $p < .05$). We present the corresponding descriptive statistics among individuals who made the selfish choice in Table 2. We present the results of

<table>
<thead>
<tr>
<th>Condition</th>
<th>Statistic</th>
<th>Ethical decision</th>
<th>Honest justification</th>
<th>Selfish justification</th>
<th>Moral identity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Control</td>
<td>$M$</td>
<td>3.25</td>
<td>3.38</td>
<td>6.29</td>
<td>5.16</td>
</tr>
<tr>
<td></td>
<td>$SD$</td>
<td>1.52</td>
<td>1.52</td>
<td>0.79</td>
<td>1.02</td>
</tr>
<tr>
<td></td>
<td>$n$</td>
<td>48</td>
<td>48</td>
<td>48</td>
<td>48</td>
</tr>
<tr>
<td>Selfish Honesty</td>
<td>$M$</td>
<td>5.23</td>
<td>5.34</td>
<td>5.28</td>
<td>5.54</td>
</tr>
<tr>
<td></td>
<td>$SD$</td>
<td>1.64</td>
<td>1.62</td>
<td>1.33</td>
<td>1.00</td>
</tr>
<tr>
<td></td>
<td>$n$</td>
<td>82</td>
<td>82</td>
<td>82</td>
<td>82</td>
</tr>
<tr>
<td>Altruistic Honesty</td>
<td>$M$</td>
<td>2.10</td>
<td>2.08</td>
<td>6.20</td>
<td>4.46</td>
</tr>
<tr>
<td></td>
<td>$SD$</td>
<td>0.96</td>
<td>1.02</td>
<td>0.97</td>
<td>1.17</td>
</tr>
<tr>
<td></td>
<td>$n$</td>
<td>40</td>
<td>40</td>
<td>40</td>
<td>40</td>
</tr>
<tr>
<td>Total</td>
<td>$M$</td>
<td>3.94</td>
<td>4.01</td>
<td>5.78</td>
<td>5.18</td>
</tr>
<tr>
<td></td>
<td>$SD$</td>
<td>1.97</td>
<td>2.00</td>
<td>1.22</td>
<td>1.13</td>
</tr>
<tr>
<td></td>
<td>$n$</td>
<td>170</td>
<td>170</td>
<td>170</td>
<td>170</td>
</tr>
</tbody>
</table>
all participants in the online supplemental materials (online supplemental materials 1.1).

Discussion

In Study 1, we found support for Hypothesis 1: individuals were more likely to behave selfishly and more likely to justify their behavior as moral when selfishness corresponded with honesty than when it simply reflected a choice to be selfish. When selfishness reflected a simple choice to prioritize one’s self interest over another’s welfare, selfishness seemed relatively unjustified (the mean rating of how “ethical” the decision was 3.25, significantly below the midpoint of the scale (4), p < .001, see Table 2). However, selfishness seemed more justified when it was associated with honesty (the mean rating of how “ethical” the decision was 5.23, significantly above the midpoint of the scale (4), p < .001, see Table 2); therefore, enabling individuals to actively choose selfish allocations. We replicated this result in a supplementary study, which included a laboratory sample and larger monetary stakes (see Study S1 in online supplemental materials 3.1. for full details). These studies provide initial evidence that honesty—a principle that is typically considered to motivate moral behavior—can enable selfish and antisocial acts.

In Study 1, we also found evidence for selective honesty (Hypothesis 2). Individuals were more likely to be honest when it helped them (was selfish) than when it harmed them (was altruistic). However, our data suggest that selective honesty was likely driven by preferences for both selfishness and honesty, rather than the propensity to use honesty as a self-serving justification. That is, the effect of honesty on selfishness was greater, but not significantly different than the effect of honesty on altruism. Therefore, we were not able to rule out the possibility that people have a genuine preference for honesty, consistent with Hypothesis 3a.

Study 2: The Effects of Honesty on Selfishness Versus Altruism in Advice-Giving Contexts

Study 2 extends our investigation by testing Hypothesis 1, 2, and 3 in a new context: advice giving. We build on paradigms used in the conflict of interest literature (e.g., Cain, Loewenstein, & Moore, 2005; Loewenstein, Cain, & Sah, 2011; Loewenstein et al., 2012) to examine how honesty influences individuals’ advice-giving behavior. Specifically, we give participants the opportunity to give truthful advice about a quantitative fact (i.e., the amount of money in a jar of coins), when doing so could be helpful or harmful.

Method

Participants. We recruited participants through two channels: advertisements at a local coffee shop (249) and MTurk (250). We used different incentives across the in-person and MTurk samples, but other than this difference, the studies were identical across samples. In total, 499 participants started the study, but six MTurk participants were kicked out of the study (before the manipulation) for failing an attention check and an additional 58 participants left the study before completing our main measures (43 from the coffee shop and 15 from MTurk). We ended up with a final sample of 435 adults (57.2% female; $M_{age} = 32$ years, $SD = 11.43$) who were eligible for analysis. For all analyses, we include all participants who responded to the relevant dependent measure. For brevity, we report our main results collapsed across samples, but we report the results split by sample in the supplement (see online supplemental materials 1.2).

We included a comprehension check at the end of the study for participants in the Selfish Honesty and Altruistic Honesty conditions. We conducted two sets of analysis—one with the full sample of participants ($N = 435$) and another omitting 42 participants who failed the check ($N = 393$). The pattern of results is qualitatively identical for both samples; all significant and nonsignificant results remain.

Procedure and materials. In Study 2, we randomly assigned participants to one of three conditions in a between-subjects design: Control, Selfish Honesty, or Altruistic Honesty. Participants completed either a “Decision task” (Control condition) or an “Advice task” (Selfish Honesty and Altruistic Honesty conditions). Across conditions, participants knew they were paired with a partner, who was a future study participant. Both the participant and their partner began with initial endowments. The endowment amounts varied based on location: On MTurk, the endowments were bonuses ranging from $0.20 to $1.00 in value, and at the coffee shop, the endowments were coupons (to be used at the coffee shop) ranging from $1.50 to $3.50 in value. The exact endowment amounts were not revealed to the participants; they simply knew the range of possible outcomes and that their decisions in the study would influence both their own and their partner’s final bonus (or coupon).

In the Decision task, Control participants learned that they would have to make a decision that would affect their own bonus and their partner’s bonus (a future participant). Then, participants simply made a decision between three options: to increase their own bonus and decrease their partner’s bonus (the selfish option), to decrease their own bonus and increase their partner’s bonus (the altruistic option), or to “make no decision,” in which case neither bonus would be affected (inaction). Participants’ choice served as our primary dependent variable.

Selfish Honesty and Altruistic Honesty participants completed the Advice task, which we modeled after existing conflict-of-interest paradigms (e.g., Loewenstein et al., 2011). Participants learned that they were assigned to the role of Adviser and would have to give advice to their partner (a future participant) about which of two jars of coins had more money in it. Participants were informed that their advice would affect their own bonus and their partner’s bonus because their partners would be asked to report which jar had more money in it “according to [their] message.” Hence, participants completing the Advice and Decision tasks had practically the same level of certainty that their decision would affect their own bonus and their partner’s bonus. As in the Decision task, participants did not know their own bonus amount or their partner’s.

Selfish Honesty and Altruistic Honesty participants learned the consequences of each message before learning which message was truthful. Specifically, participants first learned that they could

---

3 The MTurk sample chose the selfish option more often than the coffee shop sample did. Beyond shifting the distribution of choices, however, the samples did not have a qualitatively different pattern of results.
advise their partner that “Jar A has more money in it,” which would lead to their own bonus increasing and their partner’s bonus decreasing (the selfish option); advise their partner that “Jar B has money in it,” which would lead to their own bonus decreasing and their partner’s bonus increasing (the altruistic option); or “not complete the Advice task, in which case [their] bonus and [their] partner’s bonus [would] not be affected” (inaction/omission). Afterward, participants learned which jar of coins actually had more money in it. 

Selfish Honesty participants learned that Jar A had more money in it (meaning that sending the honest message was associated with the selfish option), while Altruistic Honesty participants learned that Jar B had more money in it (meaning that sending the honest message was associated with the altruistic option). Participants then decided whether to send one of the two messages or make no decision. Their choice served as our primary dependent variable.

It is worth emphasizing that we told all participants that their choices away from inaction and toward the selfish allocation were coded as 1, choosing inaction (omission) was coded as 2, and choosing a selfish allocation was coded as 3. We treated inaction (omission) and toward either the selfish or altruistic allocation. We dummy coded Selfish Honesty as 1 = Selfish Honesty condition, 0 = otherwise, and Altruistic Honesty as 1 = Altruistic Honesty condition, 0 = Otherwise.

In other words, in Study 2, we do not find support for Hypothesis 3b or Hypothesis 3c. Alternatively, we are not able to reject the effect of the Selfish Honesty condition on selfish choice to the Effect of Altruistic Honesty on altruistic choices reveals that effect sizes are not significantly different (χ² = 1.63, p = .20). In other words, in Study 2, we do not find support for Hypothesis 3b or Hypothesis 3c. Alternatively, we are not able to reject the

### Table 3

<table>
<thead>
<tr>
<th>Choice</th>
<th>Condition</th>
<th>b</th>
<th>SE</th>
<th>p</th>
<th>Lower</th>
<th>Upper</th>
</tr>
</thead>
<tbody>
<tr>
<td>Altruistic allocation</td>
<td>Altruistic Honesty</td>
<td>1.025</td>
<td>0.299</td>
<td>0.001</td>
<td>0.438</td>
<td>1.612</td>
</tr>
<tr>
<td></td>
<td>Selfish Honesty</td>
<td>0.350</td>
<td>0.365</td>
<td>0.338</td>
<td>−0.365</td>
<td>1.065</td>
</tr>
<tr>
<td></td>
<td>Constant</td>
<td>−0.738</td>
<td>0.212</td>
<td>&lt;.001</td>
<td>−1.152</td>
<td>−0.323</td>
</tr>
<tr>
<td>Selfish allocation</td>
<td>Altruistic Honesty</td>
<td>0.491</td>
<td>0.283</td>
<td>0.083</td>
<td>−0.065</td>
<td>1.046</td>
</tr>
<tr>
<td></td>
<td>Selfish Honesty</td>
<td>1.505</td>
<td>0.283</td>
<td>&lt;.001</td>
<td>0.950</td>
<td>2.059</td>
</tr>
<tr>
<td></td>
<td>Constant</td>
<td>−0.283</td>
<td>0.184</td>
<td>0.123</td>
<td>−0.643</td>
<td>0.077</td>
</tr>
</tbody>
</table>

### Post hoc tests

**Overall effect of Altruistic Honesty**

χ² = 11.74, p < .003

**Overall effect of Selfish Honesty**

χ² = 31.79, p < .001

**Effect of Altruistic Honesty on Altruism vs. Effect of Selfish Honesty on Selfishness**

χ² = 1.63, p = .201

Note. CI = confidence interval. In this multinomial regression, we ordered choices from most costly to least costly such that choosing an altruistic allocation was coded as 1, choosing inaction (omission) was coded as 2, and choosing a selfish allocation was coded as 3. We treated inaction (omission) as our base outcome. Therefore, the regression tests how Selfish Honesty and Altruistic Honesty each shifted people’s choices away from inaction (omission) and toward either the selfish or altruistic allocation. We dummy coded Selfish Honesty as 1 = Selfish Honesty condition, 0 = otherwise, and Altruistic Honesty as 1 = Altruistic Honesty condition, 0 = Otherwise.

We ran the same set of analyses in Study 2 as we had in Study 1. Table 3 and Figure 3 depict the results. Consistent with Hypothesis 1, we find that participants’ distribution of choices was significantly different in the Selfish Honesty condition and the Control condition (χ² = 31.79, p < .001). The coefficients of the multinomial logistic regression reveal that relative to the Control condition, participants in the Selfish Honesty condition shifted their choices away from inaction and toward the selfish allocation (b = 1.51, 95% CI [0.95, 2.06]).

As in Study 1, we also find that participants’ distribution of choices was significantly different in the Altruistic Honesty condition and the Control condition (χ² = 11.74, p = .003). Relative to the Control condition, participants in the Altruistic Honesty condition shifted their choices away from inaction and toward the altruistic allocation (b = 1.03, 95% CI [0.44, 1.61]).

Though having the opportunity to tell a selfish truth (Selfish Honesty condition) had a qualitatively larger effect on participants’ distribution of choices than having the opportunity to tell an altruistic truth (Altruistic Honesty condition) did, comparing the effect of the Selfish Honesty condition on selfish choice to the effect of Altruistic Honesty condition on altruistic choices reveals that effect sizes are not significantly different (χ² = 1.63, p = .20). In other words, in Study 2, we do not find support for Hypothesis 3b or Hypothesis 3c. Alternatively, we are not able to reject

---

After making their decision, participants indicated their agreement with the 10 Honesty-Humility scale items from the HEXACO-60 personality inventory (Ashton & Lee, 2009). We present the results of these measures, and how they interact with our Selfish Honesty and Altruistic Honesty conditions in online supplemental materials 1.2.
Hypothesis 3a, that honesty enables selfishness and selflessness to a similar degree.

Notably, when we examine selective honesty, we find that participants were significantly more likely to be honest in the Selfish Honesty condition than the Altruistic Honesty condition ($\chi^2 = 24.49, p < .001$). As in Study 1, though people are more likely to be honest when it helps them than when it harms them (consistent with Hypothesis 2), we cannot reject the possibility that honesty is genuinely motivating behavior (consistent with Hypothesis 3a).

Discussion

Study 2 replicates the results of Study 1 using a novel advice-giving task. Specifically, we find that although people are more likely to give honest advice when doing so is selfish rather than altruistic, this pattern of selective honesty is consistent with genuine preferences for honesty.

Study 3: Does Honesty Enable Self-Interest or Harm?

In Studies 1 and 2, we found that honesty enabled selfish behavior. However, in these studies, selfishness and interpersonal harm were confounded; participants could choose an allocation that both helped themselves and harmed another person. Therefore, it is not clear whether having the opportunity to tell the truth changed behavior by propelling people toward self-interest or interpersonal harm. Does honesty make selfishness seem more ethical, or does it make interpersonal harm seem less unethical? In Study 3, we independently manipulate selfishness and harm, and conduct mediation analyses, to gain insight into this question.

Method

Participants. We aimed to recruit 1,800 adults to participate in an online study via MTurk. There were 1,883 participants who started the study, but 143 were automatically kicked out of the study (before the manipulation) for failing an attention check, and an additional six left the survey before completing our main measures. We ended up with a final sample of 1,734 adults (1,729 of whom provided demographic information; 52.8% female; $M_{age} = 38$ years, $SD = 11.74$) who were eligible for analysis. For all analyses, we include all participants who responded to the relevant dependent measure.

Procedure and materials. In Study 3, we randomly assigned participants to a condition from a 3 (Decision context: Control, Selfish/Harmful Honesty, Selfless/Helpful Honesty) $\times$ 3 (Consequences: Self only, Other only, Self and other) between-subjects design. The Self and other condition was nearly identical to Study 1: participants were in the role of Decider and had to decide whether to do nothing and keep the allocation as is (inaction), or switch to a more selfish (Option A) or altruistic (Option B)
allocation. We varied whether Option A or Option B was associated with honesty.

However, we made two important changes. First, we no longer labeled the initial allocation as the default allocation. We simply called it the “initial allocation” to avoid suggesting that we (the experimenters) endorsed this allocation or that it should be chosen more often than Options A or B (McKenzie et al., 2006; Tannenbaum & Ditto, 2011). Second, the values associated with the different options changed. In both Studies 1 and 3, the initial allocation was $.25 for the Decider and $.25 for the Receiver. However, in Study 1 Option A was associated with $.55 for the Decider and $.05 for the Receiver, and Option B was associated with $.05 for the Decider and $.55 for the Receiver, whereas in Study 3, Option A was associated with $.45 for the Decider and $.05 for the Receiver, and Option B was associated with $.05 for the Decider and $.45 for the Receiver. We made this change so that participants could no longer use efficiency as a justification for their selfishness or honesty; the total allocation ($0.50) was constant across all three allocation options.

In the Self only condition, we changed the payouts further so that participants’ decisions only affected their own (the Decider’s) payout. Specifically, the initial allocation was $.25 for the Decider, Option A was associated with $.45 for the Decider, and Option B was associated with $.05 for the Decider. Participants knew that the Receiver would receive a randomly assigned bonus between $.05 and $.45, regardless of their decisions.5

In the Other only condition, participants’ decisions only affected their partner’s (the Receiver’s) payout. Specifically, the initial allocation was $.25 for the Receiver, Option A was associated with $.05 for the Receiver, and Option B was associated with $.45 for the Receiver. Participants knew that they (the Decider) would receive a randomly assigned bonus between $.05 and $.45, regardless of their decisions. Table 4 summarizes the nature of the three consequences conditions.

**Dependent variables.** The primary dependent variable was participants’ distribution of choices across conditions. We were particularly interested in participants’ choice of Option A (the selfish and/or harmful allocation) relative to all other choices. Participants also rated how ethical choosing Option A was, using the same Ethical Decision scale used in Study 1 ($r(1732) = .86, p < .001). Notably, in this study, we had all participants make a judgment of the same decision (Option A) "regardless of [their] actual choice.” We framed the question this way to minimize justification processes, to make clearer claims about how the choice set influenced moral judgments, and to formally test mediation.

At the end of the study, we asked participants to choose which of the outcomes (the default allocation, Option A, or Option B) was most attractive to them, as in Study 1. After participants submitted their responses, we collected demographic information, and asked participants what they thought the purpose of the study was. Participants then received a bonus payment based upon their decisions.

We conducted two main sets of analyses. First, we conducted the same multinomial logistic regressions we had conducted in Study 1 to distinguish between Hypothesis 3a, 3b, and 3c, within the Self and other condition (see Table 5). This serves as a replication of Study 1. Second, we conducted a logistic regression on the decision to select the selfish/harmful allocation, using all of our data, to examine whether honesty primarily enables self-interest or harm. Figure 4 depicts the pattern of results across conditions.

**Results**

**Behavior in the Self and Other condition: Replicating Study 1.** Consistent with Hypothesis 1, we find that participants’ distribution of choices was significantly different in the Selfish Honesty condition and the Control condition ($\chi^2 = 35.79, p < .001$). The coefficients of the multinomial logistic regression reveal that relative to the Control condition, participants in the Selfish Honesty condition shifted their choices away from inaction and toward the selfish allocation ($b = 1.27, 95% CI [0.86, 1.69]$).

As in Studies 1 and 2, we also find that participants’ distribution of choices was significantly different in the Altruistic Honesty condition and the Control condition ($\chi^2 = 24.29, p < .001$). Relative to the Control condition, participants in the Altruistic Honesty condition shifted their choices away from inaction and toward the altruistic allocation ($b = 2.40, 95% CI [1.43, 3.36]$).

Though having the opportunity to tell a selfish truth (Selfish Honesty condition) had a qualitatively larger effect on participants’ distribution of choices than having the opportunity to tell an altruistic truth (Altruistic Honesty condition) did, the effect of Altruistic Honesty on altruistic choices is actually significantly larger than the effect of Selfish Honesty on selfish choice ($\chi^2 = 4.62, p = .035$). This is likely driven by the fact that the rate of altruistic choice in the Control condition was near zero (2.5%); therefore, a moderate absolute shift in the rate of altruistic choice within the Altruistic Honesty condition (to 22.2%) represents a large effect size (i.e., a large change in the odds ratio). In summary,

---

5 We note that the there was a typo in the instructions in the Self only conditions. The first time participants saw the stimuli, the values of Option A and Option B were reversed. However, this error was corrected later in the survey. Based on the comprehension check results, this does not seem to have caused undue confusion.
we do not find support for Hypothesis 3b or Hypothesis 3c in Study 3.

Notably, when we examine selective honesty, we find that participants are significantly more likely to be honest in the Selfish Honesty condition than the Altruistic Honesty condition ($\chi^2 = 73.74, p < .001$). In other words, though people are more likely to be honest when it helps them than when it harms them (consistent with Hypothesis 2), we are not able to reject the possibility that honesty enables selfishness to a smaller degree than altruism.

Comparing all conditions: Examining whether honesty enables self-interest versus harm. In addition to replicating the analyses from Study 1 with the Self and other condition, we conducted a logistic regression on Option A ($1 = \text{selfish/harmful allocation}$, $0 = \text{all other choices}$), using all of our data. We focus on the decision to choose the selfish/harmful allocation because we are primarily interested in whether the effect of honesty on selfish harm (Hypothesis 1) is driven by the effect of honesty on selfishness or honesty on harm.

Our independent variables were Self only condition ($1 = \text{Self only condition}$, $0 = \text{other conditions}$), Other only condition ($1 = \text{Others only condition}$, $0 = \text{other conditions}$), Selfish/Harmful Honesty condition ($1 = \text{Selfish/Harmful Honesty condition}$, $0 = \text{other conditions}$), Selfless/Helpful Honesty condition ($1 = \text{Selfless/Helpful Honesty condition}$, $0 = \text{other conditions}$), Self only $\times$ Selfish/Harmful Honesty, Self only $\times$ Selfless/Helpful Honesty, Other only $\times$ Selfish/Harmful Honesty, Other only $\times$ Selfless/Helpful Honesty interactions. The Control condition of the Decision context factor, and the Self and other condition of the Consequences factor serve as the control conditions. Table 6 displays the regression results.

This analysis revealed a main effect of Other only ($p < .001$, see Table 6, Model 1, for additional statistics), such that participants were less likely to choose a selfish/harmful allocation when their choices only affected others (as opposed to both others and themselves). In other words, participants were less likely to choose a purely harmful allocation than a selfish, harmful allocation. There was also a main effect of Self only ($p < .001$), such that participants were more likely to choose a purely selfish allocation than a selfish, harmful allocation. There was also a main effect of Selfish/Harmful Honesty ($p < .001$), such that participants were more likely to choose a selfish/harmful allocation when it was honest.
These main effects were qualified by two significant interactions. There was a significant Selfish/Harmful Honesty × Self only interaction ($p < .001$), such that honesty lead to increased selfishness in the Self and other condition ($\chi^2(1, N = 399) = 36.68, p < .001$), but not in the Self only condition ($\chi^2(1, N = 373) = .018, p = .892$). In other words, a selfish, harmful allocation was more likely to be selected when it was associated with honesty than when it reflected a simple choice, but a selfish allocation (that was not associated with harm) was no more or less likely to be selected based on whether it was associated with honesty.

As shown in Figure 4 (Panel B), the majority of participants (82.3%) chose the Selfish allocation in the Control condition of the Self only condition, presumably because choosing it did not induce moral conflict or threaten one’s self-image (because being selfish did not impose a cost to another person). Therefore, adding honesty as a potential justification for choosing this allocation was unnecessary. A similar percentage of participants (82.9%) chose the Selfish allocation in the Selfish/Harmful Honesty condition of the Self only condition.

In contrast, as shown in Figure 4 (Panel A), only 35.4% chose the Selfish allocation in the Control condition of the Self and other conditions.
condition. In this condition, choosing a selfish allocation did induce moral conflict and threaten one’s self-image (because being selfish imposed a cost to another person). Therefore, participants were unlikely to choose this allocation until a moral justification, such as honesty, was available. A significantly larger percentage of participants (65.7%) chose the selfish allocation in the Selfish/Harmful Honesty condition than in the Control condition of the Self and other condition.

Furthermore, there was a significant Selfless/Helpful Honesty × Self only interaction (p < .001). As shown in Figure 4 (Panel B), significantly fewer people chose the selfish allocation in the Selfless/Helpful Honesty condition (28.1%) than in the Control condition (82.3%) of the Self only condition (χ²(1, N = 397) = 117.809, p < .001). In other words, a selfish allocation that did not harm others was chosen frequently when it reflected a simple choice, but was significantly less likely to be chosen when it was associated with lying. However, as shown in Figure 4 (Panel A), there was no difference in the frequency with which participants chose the selfish allocation in the Selfless/Helpful Honesty condition (28.1%) and in the Control condition (35.4%) of the Self and other condition (χ²(1, N = 383) = 2.31, p = .128). A selfish allocation that did harm others was chosen with relatively low frequency, regardless of whether it also involved lying.

There was not a significant Selfish/Harmful Honesty × Other only interaction (p = .491), nor was there a significant Selfless/Helpful Honesty × Other only interaction (p = .167). In other words, introducing honesty as a potential justification had similar effects in the Self and other condition and the Other only condition.

**Ethical decision.** We ran an ordinary least squares (OLS) regression on moral judgments of choosing Option A (the selfish/harmful allocation), using the same independent variables that we used in our last regression. We found a main effect of Other only (p = .006, see Table 6, Model 2, for regression coefficients and Table 7 for descriptive statistics), such that participants thought choosing a harmful allocation was less moral when it was not associated with a personal benefit than when it was. We also found a main effect of Self only (p < .001), such that participants thought choosing a selfish allocation was more moral when it was not associated with harm to others than when it was. There was also a main effect of Selfish/Harmful Honesty (p < .001), such that participants thought choosing a selfish/harmful allocation was more moral when it was associated with honesty than when it was not, and a main effect of Selfless/Helpful Honesty (p < .001), such that participants thought choosing a selfless/helpful allocation was less moral when it was associated with lying than when it was not.

These were qualified by three significant interactions. There was a significant Selfish/Harmful Honesty × Self only interaction (p < .001). Within the Self only condition, choosing the purely selfish allocation (that had no interpersonal costs) did not seem more moral when it was associated with honesty (M = 5.55, SD = 1.66) than when it was not (M = 5.64, SD = 1.51; t = .48, p = .634). Selfishness was seen as relatively moral, regardless of whether it was associated with honesty (5.64 is significantly greater than the midpoint of the scale [4], p < .001). However, within the Self and other condition, choosing the selfish allocation (that did have interpersonal costs) did seem more moral when it was associated with honesty (M = 5.24, SD = 1.70) than when it was not (M = 3.63, SD = 1.92; t = −9.12, p < .001).

There was also a significant Selfish/Harmful Honesty × Other only interaction (p = .022). Within the Other only condition, choosing the purely harmful allocation (that had no selfish benefits) seemed more moral when it was associated with honesty (M = 5.32, SD = 1.58) than when it was not (M = 3.15, SD = 1.97; t = −12.24, p < .001) and the magnitude of this effect was greater than the effect of honesty within the Self and other condition (t = −9.12, p < .001).

Furthermore, there was a significant Selfless/Helpful Honesty × Self only interaction (p < .001). Within the Self only condition, choosing the purely selfish allocation (that had no interpersonal costs) seemed less moral when it was associated with a lie (M = 3.30, SD = 1.99) than when it was not (5.64, SD = 1.51; t = 13.25, p < .001). Within the Self and other condition, choosing the selfish allocation (that was associated with interpersonal costs) also seemed less moral when it was associated with a lie (M = 3.01, SD = 1.62) than when it was not (M = 3.63, SD = 1.92; t = 3.46, p = .001), though this effect was significantly smaller in magnitude than in the Self only condition because the selfish allocation was already seen as relatively immoral.
There was not a significant Selfless/Helpful Honesty × Other only interaction ($p = .173$). Within the Other only condition, choosing the purely harmful allocation (that had no selfish benefits) did not seem less moral when it was associated with a lie ($M = 2.87, SD = 1.77$) than when it was not ($M = 3.15, SD = 1.97; t = 1.55, $p = .121$). Although within the Self and other condition, choosing the selfish allocation (that was associated with interpersonal costs) did seem less moral when it was associated with a lie than when it was not ($t = 3.46, p = .001$), the magnitude of this effect was not significantly different than that in the Other only condition.

**Mediation analyses.** We conducted exploratory mediation analyses to examine whether participants’ decisions to choose Option A (the selfish/harmful allocation) were driven by perceptions of Option A’s ethicality within each condition. Specifically, we ran a series of six mediation models. In every model, the mediator was the perceived ethicality of choosing Option A (Ethical choice) and the dependent variable was the participant’s decision to choose Option A (1 = Option A (Ethical choice) and the dependent variable was the participant’s decision to choose Option A (1 = Option A, 0 = Control condition). In these models, the independent variable was the Selfish/Harmful Honesty condition ($1 = Selfish/Harmful Honesty condition, 0 = Control condition; the Selfless/Helpful Honesty condition was necessarily omitted). Model 1 examined the Other only condition, Model 2 examined the Self only condition, and Model 3 examined the Self and other condition. Models 4–6 examined how the Selfless/Helpful Honesty choice set influenced perceptions of the ethicality of Option A and, thus, the choice of Option A, relative to the Control condition. In these models, the independent variable was the Selfish/Harmful Honesty condition ($1 = Selfish/Harmful Honesty condition, 0 = Control condition; the Selfless/Helpful Honesty condition was necessarily omitted). Model 4 examined the Other only condition, Model 5 examined the Self only condition, and Model 6 examined the Self and other condition. All models were run using the PROCESS Macro for SPSS (Model 4, 10,000 samples; Hayes, 2017). Table 8 depicts the results.

Model 1 reveals that when participants’ choices only affected themselves, Option A (the purely selfish option) did not seem more moral when it was associated with honesty than when it was not and, therefore, choices did not differ between these two conditions.

### Table 8

**Mediation Analyses in Study 3**

<table>
<thead>
<tr>
<th>Consequence condition</th>
<th>IV: Control versus Selfish/Harmful Honesty</th>
<th>IV: Control versus Selfless/Helpful Honesty</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Model</td>
<td>Indirect effect [95% CI]</td>
</tr>
<tr>
<td>Self only</td>
<td>1</td>
<td>$-0.3 [-0.15, -0.08]^b$</td>
</tr>
<tr>
<td>Others only</td>
<td>2</td>
<td>$0.67 [0.32, 1.07]^b$</td>
</tr>
<tr>
<td>Self and others</td>
<td>3</td>
<td>$-0.64 [-0.41, -0.94]^b$</td>
</tr>
</tbody>
</table>

**Note.** CI = confidence interval. Data was split by consequence condition, and mediation analyses were performed using the PROCESS macro for SPSS, model 4, with 10,000 samples. Indirect effects that have different superscript letters within each column significantly differ from each other.
However, when participants’ choices only affected others, Option A (the purely harmful option) seemed more moral when it was associated with honesty than when it was not, which led participants to choose it more often (Model 3). Similarly, when participants’ choices affected others and themselves, Option A (the harmful, selfish option) seemed more moral when it was associated with honesty than when it was not, which led participants to choose it more often (Model 3).

Model 4 reveals that when participants’ choices only affected themselves, Option A (the purely selfish option) also seemed less more moral when it was associated with deception than when it was not, which led participants to choose it less often. Similarly, when participants’ choices affected others and themselves, Option A (the harmful, selfish option) seemed less moral when it was associated with deception than when it was not, which led participants to choose it less often (Model 6). However, when participants’ choices only affected others, Option A (the purely harmful option) seemed no less moral when it was with a lie than when it was not and, therefore, choices did not differ between these two conditions (Model 5).

**Discussion**

Study 3 examines how honesty influences choices that only affect oneself, only affect others, or affect both oneself and others. Our results suggest that honesty helps people adjudicate between the desire to help oneself and the desire to avoid harm to others. When selfishness did not come at a cost to others, participants did not have moral reservations about acting selfishly. Therefore, associating selfishness with honesty did not change moral judgments of selfishness, nor did it change choice; the majority of participants (over 82%) chose the selfish allocation in both conditions. However, when selfishness did come at a cost to others, selfishness seemed relatively unethical.

When selfishness was associated with interpersonal harm, the majority of participants chose to do nothing (62.1%) rather than actively help themselves or others. However, associating selfish, interpersonal harm with a moral justification, such as honesty, increased the perceived ethicality of acting selfishly and, therefore, pushed participants away from inaction and toward selfishness.

Even when interpersonal harm was not beneficial for oneself, associating harm with honesty increased its perceived morality and, therefore, increased participants’ likelihood of harming others. Taken together, these results suggest that honesty enables selfish harm primarily by reducing the moral cost of harming others, rather than making selfishness itself more justified. These results also provide further evidence that people have genuine preferences for honesty (consistent with Hypothesis 3a). People are more likely to pursue (both helpful and harmful) actions when those actions are associated with honesty, even when such actions have no benefits for the self.

**Study 4: How Does Honesty Influence Information Avoidance?**

Study 3 suggests that honesty reduces the moral cost of harming others. In Study 4, we document a strategy that allows people to circumvent the moral cost of honest harm altogether: information avoidance. Knowing that one’s honesty causes interpersonal harm puts people in an ethical dilemma. One way that individuals can avoid this dilemma is by ignoring information about the harm that honesty might cause. In Study 4, we test this possibility, thereby testing Hypothesis 4.

**Method**

**Participants.** We made the a priori decision to run this study for one laboratory session (1 day in a computer lab, which typically recruits 100–200 participants) at a university laboratory in Philadelphia, PA and then end data collection. There were 223 participants who started the survey, but 73 participants were automatically kicked out of the study (before the manipulation) for failing an attention check. We ended up with 150 adults (67% female; M_age = 23 years, SD = 7.08) who completed the study and were eligible for analysis.

**Procedure and materials.** Participants played an economic game similar to the one we used in Studies 1 and 3, and were assigned to the role of Decider. Participants learned that one participant would be randomly selected to receive the amount of money associated with the decisions he or she made as the Decider in this study.

All participants started by learning that they would have the opportunity to choose an allocation for themselves (the Decider) and their partner (the Receiver). We randomly assigned participants to one of three conditions (Control, Selfish Honesty, Selfless Honesty) in a between-subjects design.

In the Control condition, participants knew that the two options were:

- Option A: $7 for the Decider and SX for the Receiver
- Option B: $3 for the Decider and SY for the Receiver

Participants did not know the numbers associated with X or Y. However, we gave participants three possible decisions: (a) Choose Option A (without finding out X), (b) Choose Option B (without finding out Y), or (c) Find out the values of X and Y and then make a decision. Therefore, participants in the Control condition had the opportunity to either (a) choose a larger allocation without knowing how it affected their partner (Option A), (b) choose a smaller allocation without knowing how it affected their partner (Option B), or (c) find out how their allocations affected their partner before making a choice.

In the Selfish Honesty and Selfless Honesty conditions, we told participants that their allocation depended on their choice and the outcome of a random number generator. As in the previous studies, we told participants that a random number generator would generate a number between 1 and 9. Participants then had to report whether the number was ODD or EVEN. If participants reported that the number was ODD, the allocation would switch to Option A ($7 for the Decider and SX for the Receiver); if they reported that the number was EVEN, the allocation would switch to Option B ($3 for the Decider and SY for the Receiver). Participants could also choose not to report a number, in which case they would find out the values associated with X and Y before making an allocation decision.

Participants then saw that the random number generator produced either an odd number (Selfish Honesty) or an even number (Selfless Honesty). Therefore, participants in the Selfish Honesty condition had the option to either (a) honestly report...
that the number was odd and receive a large allocation, without knowing how it affected their partner (Option A), (b) dishonestly report that the number was even and receive a smaller allocation, without knowing how it affected their partner (Option B), or (c) choose not to report a number so that they could find out how the allocations affected their partner. Participants who chose to find out how the allocations affected their partner then made a choice between Option A and Option B with full information.

Participants in the Selfless Honesty condition had the opportunity to either (a) dishonestly report that the number was odd and receive a large allocation, without knowing how it affected their partner (Option A), (b) honestly report that the number was even and receive a smaller allocation, without knowing how it affected their partner (Option B), or (c) choose not to report a number so that they could find out how the allocations affected their partner. Participants who chose to find out how the allocations affected their partner then made a choice between Option A and Option B with full information.

Participants in the Selfless Honesty condition had the opportunity to either (a) dishonestly report that the number was odd and receive a large allocation, without knowing how it affected their partner (Option A), (b) honestly report that the number was even and receive a smaller allocation, without knowing how it affected their partner (Option B), or (c) choose not to report a number so that they could find out how the allocations affected their partner. All participants then made a choice. This served as our primary dependent variable. Participants who chose to find out the values associated with X and Y then learned that $X = \$2$ and $Y = \$4$ and made a final allocation choice (and they no longer had to report a number). Participants then provided demographic information and were dismissed. We randomly selected one participant to receive a bonus payment based on their decision, and paid this bonus 1 week after the study concluded.

We ran similar analyses in Study 4 to those we had run in Studies 1 and 2. However, logistic regression was not suitable because the frequency of selfless behavior in some conditions was 0% (King & Zeng, 2001). Penalized Maximum Likelihood Estimation maintains the assumptions of logistic regression, but corrects for rare events (Firth, 1993). Therefore, we used this method to analyze our data (using the firthlogit command in Stata).

Specifically, we used firth logistic regression to test how Selfish Honesty and Selfless Honesty influenced participants’ choices of the selfish (high payoff) and selfless (low payoff) allocation, respectively, relative to the Control condition. We complement these analyses with chi-square tests of proportions, which allow us to examine how the full distribution of choices varied across each condition. Figure 5 displays these results.

Results

Selfish Honesty versus Control. Participants’ distribution of choices was significantly different in the Selfish Honesty condition than the Control condition ($\chi^2 = 13.69, p < .001$). Specifically, participants were significantly more likely to choose the selfish (high payoff) allocation, without finding out how their decision affected others, in the Selfish Honesty condition relative to the Control condition (74.5% vs. 38.0%) and significantly less likely to omit information so they could find out how their choice affected others (25.5% vs. 62.0%). No participants chose the selfless (low payoff) allocation in either condition. A firth logistic regression (in which we dummy coded Selfish Honesty as 1 = Selfish Honesty condition, 0 = Control condition, and coded

---

6 In this study, participants did not necessarily know whether they would still need to report whether the number was ODD or EVEN to choose an allocation (after finding out X and Y), or whether they could just choose an allocation. We removed this ambiguity in the subsequent study (Study 5).
choice as Selfish allocation (high payoff) = 1, 0 = Otherwise) confirmed that there was a significant effect of Selfish Honesty on selfish allocation choice ($b = 1.53$, $SE = .43$, $p < .001$, 95% CI [.69, 2.37]).

**Selfless Honesty versus Control.** Participants’ distribution of choices was also significantly different in the Selfless Honesty condition relative to the Control condition ($\chi^2 = 13.93$, $p = .001$). Specifically, participants were significantly more likely to choose the selfless (low payoff) allocation, without finding out how their decision affected others, in the Selfless Honesty condition relative to the Control condition ($24.5\%$ vs. $0.0\%$; $\chi^2 = 13.93$, $p = .001$). Participants were no more or less likely to find out how their choice affected others ($46.9\%$ vs. $62.0\%$; $\chi^2 = 2.26$, $p = .132$) or to choose the selfish (high payoff) allocation ($28.6\%$ vs. $38.0\%$; $\chi^2 = 9.90$, $p = .320$) across conditions. A firth logistic regression (in which we dummy coded Selfless Honesty as 1 = Selfless Honesty condition, 0 = Control condition, and coded choice as selfless allocation (low payoff) = 1, 0 = Otherwise) confirmed that there was a significant effect of Selfless Honesty on selfish allocation choice ($b = 3.52$, $SE = .146$, $p = .016$, 95% CI [.66, 6.37]). The highly overlapping confidence intervals for the effect of Selfless Honesty on selfish allocation choice and the effect of Selfless Honesty on selfless allocation choice are consistent with Hypothesis 3a: honesty has a similar effect on both selfish and selfless behavior, suggesting that it serves as a genuine motivator of behavior.

Notably, when we examine selective honesty, we find that participants were significantly more likely to be honest in the Selfless Honesty condition than the Selfless Honesty condition ($\chi^2 = 25.01$, $p < .001$). As in our previous studies, though people are more likely to be honest when it helps them than when it harms them (consistent with Hypothesis 2), we cannot reject the possibility that honesty is genuinely motivating behavior (consistent with Hypothesis 3a).

There were no differences in final allocation choices among the participants who selected to find out the values of X and Y, $\chi^2(2, N = 150) = 2.13$, $p = .34$ (Control: 93.5% chose Option A, the high payoff choice; High Payoff Honesty: 100% were honest and chose Option A; Low Payoff Honesty: 87.0% chose Option A).

### Discussion

Study 4 provides support for Hypothesis 4: individuals are more likely to avoid learning the consequences of their actions when those actions are honest. Participants were less likely to consider how their choice of allocation would affect their partner when the allocation was associated with honesty than when the allocation simply reflected a personal choice. As in Studies 1–3, we find that honesty had similar magnitude effects on both selfish and selfless choices. Therefore, the effects of honesty on information avoidance are more consistent with genuine preferences for honesty (supporting Hypothesis 3a), rather than with motivated processes.

### Study 5: How Does Honesty Influence Information Avoidance in Absence of Self-Interest?

In Study 5, we test whether honesty can cause social harm (vis-à-vis information avoidance), even when actors have no personal incentive to be honest or dishonest. We also return to the advice-giving paradigm we used in Study 2, to test whether information avoidance persists in the advice-giving context.

### Method

**Participants.** We aimed to recruit 200 adults to participate in an online study via Academic Prolific (a web-based research survey platform), based on the a priori goal of recruiting 100 participants per condition. There were 213 participants who started the study, but 24 were automatically kicked out of the study (before the manipulation) for failing an attention check, and an additional nine left the survey before completing our main measures. We ended up with a final sample of 180 adults who passed the comprehension checks, completed the entire study, and were eligible for analysis (55% female; $M_{age} = 28$ years, $SD = 9.64$).

**Procedure and materials.** In this study, we randomly assigned participants to one of two conditions in a two cell (Honest Motive vs. No Honest Motive) between-subjects design. Participants completed an Advice task, which was adapted from the paradigm we used in Study 2. We varied whether participants honestly knew the amount of money in the jar of coins.

Participants in the No Honest Motive conditions saw five different jars of coins with varying amounts ($3, $18, $23, $29$, and $50) and were told that their partner would see one of those jars selected at random. Therefore, participants could not use honesty to motivate their behavior because they did not have any insight into what truthful advice would look like. Alternatively, participants in the Honest Motive conditions saw the jar of coins that their partner would see and were told that the jar contained exactly $3. Therefore, participants who were motivated by honesty could simply advise their partner that there were $3 in the jar.

All participants were told that their partner’s bonus payment would depend on their partner’s own guess about the amount of money in the jar that they would see. However, they did not know exactly what determined their partner’s payment. Participants simply knew that their partner “could receive a bonus payment or a penalty for guessing the correct amount of money in the jar that they see, or guessing a specific amount.” We obfuscated how their partner’s payment was determined because we were interested in whether participants would choose to learn more about it before sending their message. Participants own bonus payment did not depend on their partner’s guess; instead they would get a fixed $0.10 bonus.

**Dependent variables.** After reading the information about the task and passing a comprehension check, participants decided which message to send their partner or whether they would like to learn what determines their partner’s bonus payment before choosing a message. Participants had the option to send one of five messages, each indicating that the jar contained a specific dollar amount with the values of $3, $18, $23, $29, or $50, or the option to “find out what determines my partner’s bonus payment before deciding what message to send.” Participants knew that if they opted to find out their partner’s payment scheme, they would then have to choose one of the five aforementioned messages. Therefore, finding out the payment structure could present participants with a dilemma between telling a harmful truth and telling a helpful lie. After making their choice, all participants who opted to find out their partner’s payment learned that their partner would...
“receive $1 if he or she guesses that there are $3 in the jar of coins” and then chose one of the five messages.

The choice to learn what determines their partner’s payment was our primary dependent variable, as indicated in our preregistration. Additionally, we were interested in which message participants chose to send as their initial decision (if they chose to send a message).

After choosing a message, participants rated their agreement with a series of statements about the importance of truth and how they justified their choice. For these measures we used 7-point Likert scale anchored at 1 = strongly disagree and 7 = strongly agree. We measured participants’ agreement with the importance of truth using three items (α = .76): “I care deeply about being honest,” “Speaking the truth is the most important value,” and “I always tell the truth.” We also measured participants’ justification for their choice of what message to send. Participants rated their agreement with three distinct justification items: “I chose the easiest course of action” (Easy Choice Justification), “I made my decision because I wanted to avoid an ethical dilemma” (Avoid Dilemma Justification), and “I made my choice because it was honest” (Honest Justification).

At the end of the study, we collected demographic information. Participants received their base payment at the conclusion of the study and, within 2 weeks, received their appropriate bonus payment.

Results

Choice to look. Whether one could motivate their behavior with honesty significantly influenced the choice to look at the social consequences of one’s words, χ²(1, N = 180) = 5.29, p = .021. Similar to Study 4, participants were significantly less likely to look at the consequences of their words in the Honest Motive condition (52.6%) than in the No Honest Motive condition (69.4%).

Those who did not look at the consequences of their messages primarily sent the honest message (“There are $3 in the jar of coins,” 45.3%) in the Honest Motive condition, and sent a variety of messages in the No Honest Motive condition (3.5%, 2.4%, 7.1%, 8.2%, and 9.4%) sent the message, “There are $3 [18, 23, 29, 50] in the jar of coins,” respectively.

Attitudinal measures. We report the descriptive statistics of all attitudinal measures in Table 9.

Importance of truth. A one-way analysis of variance (ANOVA) revealed no main effect of Honest Motive, F(1, 179) = .19, p = .66, η² = .001.

Justifications. We also conducted one-way ANOVAs on each self-reported justification measure. Participants in the Honest Motive condition were significantly more likely to justify their decision with honesty, F(1, 179) = 18.40, p < .001, η² = .094, and marginally more likely to justify their decision with their desire to avoid an ethical dilemma, F(1, 179) = 3.48, p = .064, η² = .019, than participants in the No Honest Motive condition. We find no effect of condition on the Easy Choice Justification, F(1, 179) = .09, p = .76 η² = .001.

Discussion

Study 5 provides evidence that advisors are less likely to look at the social consequences of their advice when they know their advice is honest. We find that this effect persists even when individuals have no personal incentive to give honest advice, suggesting that people have a genuine (nonmotivated) preference for honesty (Hypothesis 3a). Consistent with this account, we find that decision-makers believe that their information avoidance is driven by a desire to be honest, rather than a desire to avoid ethical dilemmas or make simple (less cognitively demanding) choices. We interpret these justifications with caution, however, given that participants’ self-reports may reflect their own desire to preserve their self-image.

We also note that curiosity may explain why participants sought out (or did not seek out) information about the social consequences of their actions. Nonetheless, honesty seemed to have minimized curiosity, causing people to ignore the social consequences of their actions more when honesty was relevant to one’s choice than when honesty was irrelevant. We encourage future scholars to examine how moral principles can influence curiosity, particularly as it relates to seeking out full information about the externalities of one’s actions.

General Discussion

Across five experiments, we explore when, why, and how honesty enables interpersonal harm. In Study 1, we find that honesty causes individuals to choose harmful and selfish options, even when doing so is unnecessary (i.e., when omission was possible). In other words, honesty enables selfish harm, consistent with Hypothesis 1. We also find that people are selectively honest. That is, people are more likely to be honest when honesty is associated with selfish harm than altruism, consistent with Hypothesis 2. We also explore whether genuine or motivated preferences for honesty underlie these effects. To do so, we compare the

<table>
<thead>
<tr>
<th>Condition</th>
<th>Statistic</th>
<th>Importance of truth</th>
<th>Honest justification</th>
<th>Avoid dilemma justification</th>
<th>Easy decision justification</th>
</tr>
</thead>
<tbody>
<tr>
<td>Honest motive</td>
<td>M</td>
<td>5.59</td>
<td>6.26</td>
<td>4.33</td>
<td>4.04</td>
</tr>
<tr>
<td></td>
<td>SD</td>
<td>0.91</td>
<td>0.91</td>
<td>1.86</td>
<td>1.73</td>
</tr>
<tr>
<td></td>
<td>n</td>
<td>95</td>
<td>95</td>
<td>95</td>
<td>95</td>
</tr>
<tr>
<td>No honest motive</td>
<td>M</td>
<td>5.53</td>
<td>5.53</td>
<td>3.81</td>
<td>3.96</td>
</tr>
<tr>
<td></td>
<td>SD</td>
<td>0.92</td>
<td>1.36</td>
<td>1.84</td>
<td>1.64</td>
</tr>
<tr>
<td></td>
<td>n</td>
<td>85</td>
<td>85</td>
<td>85</td>
<td>85</td>
</tr>
</tbody>
</table>
effect of honesty on selfish harm to the effect of honesty on altruism and find that honesty enables selfish harm and altruism to similar degrees. We find that selective honesty is more consistent with independent preferences for selfishness and honesty, rather than the motivated use of honesty, consistent with Hypothesis 3a. We replicate this pattern of results in Study 2, in which we examine honesty in the context of giving advice.

In Study 3, we explore how honesty independently influences selfishness and harm. Our results suggest that honesty increases selfish harm primarily by making harm seem less immoral, rather than making selfishness seem more moral. Selfish behavior that is not associated with interpersonal harm is not deemed immoral, and adding honesty as a potential justification to this behavior does not meaningfully influence perceptions of its immorality. As a result, people choose selfish, nonharmful allocations quite frequently, regardless of whether doing so corresponds with a moral principle (in this context, the principle of honesty). In contrast, selfish harmful behavior is deemed immoral. As a result, people typically choose to do nothing (inaction) rather than actively choose selfish, harmful allocations. However, when selfish, harmful behavior is associated with honesty it is perceived to be morally justified. Therefore, adding honesty as a potential justification for selfish harmful behavior increases its frequency. We also find that adding honesty as a potential justification for nonselfish harmful behavior also makes this behavior seem more justified, thereby enabling interpersonal harm, independent of any gains to the self. These findings provide further evidence that honesty genuinely motivates behavior (Hypothesis 3a); honesty enables behavior that participants have no personal interest in pursuing.

By examining the rates of different behaviors across conditions in Study 3 (see Figure 4), we can see how honesty, and moral principles more broadly, helps people resolve moral conflicts. People typically feel conflicted when they have to choose between being selfish and violating a moral principle. Therefore, they often prefer to avoid choices that force this tradeoff (by engaging in inaction). However, adding a moral justification to selfish behavior helps people resolve the conflict and, thus, shifts people away from inaction and toward the selfish action. To illustrate this psychology, note that in Study 3 the plurality of people chose inaction when they faced a conflict between being selfish and harming others (in the Control-Self and other condition) and between being selfish and lying (in the Altruistic Honesty-Self and other and the Selfless Honesty-Self only conditions). However, the majority of people chose Selfish Honesty when selfishness and honesty were aligned (in the Selfish/Harmful Honesty-Self and other and the Selfish Honesty-Self only conditions).

Studies 4 and 5 build on these findings by exploring another way in which honesty can help people resolve moral conflicts, by increasing information avoidance. Whereas in Study 3 we found that adding honesty as a justification for behavior reduced the perceived moral cost of harming others, in Studies 4 and 5 we found that honesty licenses people to avoid finding out about the harm their behavior causes in the first place. In Study 4, people were more likely to make (both selfish and selfless) decisions without finding out the social welfare implications of their decisions, when their decisions were associated with honesty. In Study 5, we found that these information-avoidance effects are independent of self-interest, thereby providing further support for Hypothesis 3a: honesty serves as a genuine motivator of behavior, rather than a self-serving justification for behavior.

These findings make a number of theoretical, methodological, and practical contributions to our understanding of ethical decision-making and honest communication. First, we expand our basic understanding of when and why people are honest by studying the conditions under which people engage in honesty. We find that although people are more likely to be honest when it helps them than when it hurts them, honesty seems to exert a genuine (i.e., unconditional) force on both selfish and selfless behavior. By studying the selective use of moral principles in the context of honesty, we also gain insight into the role that inaction plays in influencing moral choices. We find that people tend to favor doing nothing (remaining silent) when honesty is personally costly. However, when honesty does not come at a cost to the self (e.g., when it is selfish, or when it is helpful to others but not costly), people choose to be honest at very high rates. Practically, this suggests that truths are most likely to surface when they are not costly to the communicator. Theoretically, this suggests that honesty, and moral principles broadly, motivate behavior (relative to inaction) in the presence of moral conflict.

Second, we introduce new methodologies for testing the mechanisms that underlie selective morality. Specifically, we clarify the importance of including control conditions for studying the nature of selective morality and outline a set of empirical tests (see Figure 1) that allow researchers to more carefully examine whether a moral principle is serving as a justification versus a genuine motivator of behavior.

Third, we deepen our insights into how people manage their moral self-image. Specifically, we demonstrate that individuals actively manage and constrain their choice sets to simplify moral decision-making. For example, in Studies 4 and 5, participants chose not to learn more about the consequences of one of their options so that they would not face an ethical dilemma. If they had learned that the selfishly honest option also harmed others, they would have been less likely to pursue it, or at least felt guilty about pursuing it.

The propensity to look away from the harm caused by moral actions also contributes to our understanding of the perceptual link between morality and harm. Judgments of harm and morality are inextricably linked (Gray et al., 2014; Schein & Gray, 2016, 2018); when people view actions as harmful, they naturally view them as wrong, and people think of immoral acts as typically harmful (Gray, Waytz, & Young, 2012). However, there are numerous quotidian situations in which morally justified actions cause harm—for example, when individuals share honest, but hurtful opinions or when they report others’ transgressions. Our research suggests that individuals may choose to look away from the social harm caused in these types of situations. As a result, individuals may rarely grapple with the idea that their moral actions cause harm; thereby, keeping the harm-morality link intact.

By documenting the prevalence of information avoidance in the context of ethical dilemmas, we also document novel negative downstream consequences of moral principles. Though past scholars have noted (Atran & Ginges, 2012; Janoff-Bulman & Carnes,
2013; Neuberg et al., 2014; Rai & Fiske, 2011) and studied (Fiske & Rai, 2014; Gómez et al., 2017; Rai, Valdesolo, & Graham, 2017) how moral motivations can lead to harm, existing literature has largely failed to disentangle the causal effect of any given moral principle on harmful and helpful action. The present research fills this gap, demonstrating that simply linking an action with honesty can enable harm.

Limitations and Future Directions

As with any research program, the current set of studies must be interpreted in light of several limitations that can be addressed with future research.

Further understanding genuine versus motivated preferences for honesty. First, it is important to note that although our data are more consistent with genuine than motivated preferences for honesty, we certainly believe there are contexts in which people do use honesty as a self-serving justification. In Studies 1 and 2, for example, honesty enabled selfish harm more than altruism, though these effects were not significantly different. It is possible that with great enough power, or stronger manipulations, we would see significant evidence that honesty enables selfishness more than altruism, consistent with a motivated account of honesty.

In Studies 3, 4, and 5, however, honesty did not enable selfish harm more than altruism, and these studies also provide evidence that honesty enables harm, independent of self-interest. These studies provide the clearest evidence that genuine preferences for honesty certainly can enable socially harmful behavior, independent of any motivated processes.

It is possible that we do not find motivated preferences for honesty, in part, because we gave participants relatively little “moral wiggle room” to construe honesty flexibly. Although our experiments afforded us experimental control they also limited participants’ choice sets relative to the choices, communicators face in everyday conversation. Specifically, participants in our studies could simply lie, omit information, or tell the truth, whereas communicators outside of the laboratory face a large range of tactics that vary in honesty (Levine, Roberts, & Cohen, 2020). It is possible that communicators only engage in complete, unvarnished truth-telling when it benefits them and, therefore, that complete honesty in conversation is often motivated by self-serving preferences. Furthermore, it is possible that people’s beliefs about what opinions and statements are true are malleable, such that honesty serves as a post hoc justification for hurtful communication. An initial pilot study (see online supplemental materials 2 for details) suggests this may be the case: 66.5% of people were able to recall at least one instance in which someone they knew used honesty as a justification for harming others. We encourage future scholars to examine the nature of selective honesty outside the confines of the lab.

Extensions to different values. We are also interested in whether honesty functions differently than other moral values. We demonstrate that honesty can enable harm, even in the absence of self-interest, but we expect that any moral principle—or policy, more broadly—could have similar effects. For example, a commitment to justice may enable decision-makers to punish offenders without considering the full consequences of their actions (e.g., how punishment would affect offenders’ families or society).

How choice sets influence genuine versus self-serving honesty. Important future work is also needed to understand how the choices available to a communicator influence a communicator’s willingness to trade off different values, and their subsequent likelihood of being honest. Across our studies, we include inaction and omission options in participants’ choice sets. While this is an important feature of real-world decisions—in most situations, people can exit a situation, choose the status quo, or stay silent—this feature may influence people’s propensity to engage in selective honesty. We posit that inaction, or omission, options can serve as a moral compromise when acting morally is personally costly. As a result, people choose omission when honesty is costly, but are compelled to speak up when honesty is selfish. Although we find evidence for such a pattern of results in our data, it is not clear whether omission options are necessary for selective honesty to emerge. We ran two supplementary studies to explore this question (see Studies S3 and S4 in online supplemental materials 3.3–3.4) and found that patterns of selective honesty were somewhat stronger when omission was possible than when omission was not possible, but these differences were not significant. Future, higher-powered studies are needed to precisely test whether omission options causally influence selective honesty.

How individual differences influence genuine versus self-serving honesty. Examining individual differences in the propensity to engage in selective honesty could also be revealing. In Study 2 we ran exploratory analyses to examine whether Honesty-Humility, a personality trait that captures individual differences in honesty, sincerity, and generosity (Ashton et al., 2004; Ashton & Lee, 2008; Lee & Ashton, 2008) moderated the selective use of honesty (see online supplemental materials 1.d for details). We found that low Honesty-Humility individuals chose to be honest more often when it helped them than when it harmed them, whereas high Honesty-Humility individuals chose to be honest regardless of whether it helped or harmed them. These results highlight the importance of accounting for individual differences to precisely understand how incentives shift behavior across populations. We encourage future research to examine the role that other individual differences may play in the pathway from honesty to help and harm.

Judgments of morally motivated information avoidance. Future work should also examine how observers judge others who follow moral rules without looking at their consequences. On the one hand, consistently following moral rules without considering their interpersonal consequences in any particular case may signal integrity. On the other hand, choosing not to seek out information about doing harm may signal low compassion. Just as recent research has looked at the consequences of uncalculating cooperation (i.e., cooperating without looking at the personal costs of doing so; Hoffman et al., 2015; Jordan et al., 2016), we encourage future research to examine the consequences of uncalculating morality (i.e., following moral principles without looking at the social costs of doing so).

Actor-target asymmetries in judgments of harmful honesty. Finally, it will be interesting to examine whether communicators make fundamentally different judgments of their harmful honesty than do targets and observers. Our results suggest that communicators’ decisions to engage in harmful honesty reflect genuine,
rather than self-serving motives. However, other parties may be motivated to derogate the intentions of actors, and may assume that harmful honesty reflects pernicious rather than admirable motives.

We ran two exploratory studies to begin to address this question (see Studies S4 and S5 in the online supplemental materials). In Studies S4 and S5, we explored how communicators judge their selfish, honest acts relative to how targets and observers judge them. In Study S4, we find that communicators’ make more favorable attributions of selfish, harmful honesty than targets do, when communicators judge their own actions. However, these effects are somewhat weaker when communicators judge a randomly assigned action, rather than their own decision (Study S5). In Study S5, we also find that communicators who imagine engaging in harmful honesty judge their moral identity to be higher than observers believe it to be, but that these asymmetric judgments do not extend to other perceptions (such as beliefs about the ethicality of the decision). Taken together, these results suggest that communicators do likely have more positive impressions of their harmful honest acts than other parties do, but more research is needed to understand the boundaries of these effects.

Conclusion

Moral principles, like honesty, equality, and justice, are assumed to be positive forces that enable good behavior and constrain bad behavior. The present research challenges this assumption by demonstrating that—even absent of motivated processes—moral principles can enable selfish and interpersonally harmful behavior. People are more likely to act selfishly, cause harm, and ignore the social consequences of their actions when these behaviors are associated with honesty than when they are not.

References


E-Mail Notification of Your Latest Issue Online!

Would you like to know when the next issue of your favorite APA journal will be available online? This service is now available to you. Sign up at https://my.apa.org/portal/alerts/ and you will be notified by e-mail when issues of interest to you become available!