No influence of simple moral awareness cues on cheating behaviour in an online experiment

Luu Duc Toan Huynh Philipp Stratmann Rainer Michael Rilke

November 10, 2023

Abstract

We study the influence of moral awareness cues on cheating in an online experiment (n=551). People's awareness of ethical issues is a pre-condition of moral behavior. The results show that reminding people of the different ethical dimensions of their actions does not reduce cheating. Our results highlight that raising moral awareness in an online experiment is not sufficient to mitigate cheating behavior.

Keywords: Cheating; moral awareness; behavioral ethics
JEL Classification: C91; J30; D91
PsycInfo classification: 3120; 3450

Declaration of Interest statement: The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Luu Duc Toan Huynh (Corresponding author, Queen Mary University of London, United Kingdom), Philipp Stratmann (WHU - Otto Beisheim School of Management, Germany), and Rainer Michael Rilke (WHU - Otto Beisheim School of Management, Germany).

Highlights

- We study the influence of moral awareness cues on cheating in an online experiment.
- We design cues based on Rest's (1986) theory of moral awareness.
- Moral awareness cues did not reduce incidents of cheating.

1 Introduction

The dilemma of whether to behave truthfully or to cheat in certain situations is ubiquitous in individuals and societies in general. In response to a growing body of research on cheating behaviour, economists have been exploring its determinants to examine how the severity of the consequences of telling a lie can be mitigated. One of the most robust findings in the literature on cheating is that individuals are more honest than conventional economic theory, which assumes income maximisation preferences, would suggest (Abeler et al., 2019; Leib et al., 2021). Several theories have tried to explain why individuals deviate from the standard economic model when facing moral dilemmas. For example, the morality preference hypothesis (Tappin and Capraro, 2018; Capraro et al., 2022; Capraro and Perc, 2021) assumes that acts of kindness and honesty, specifically, are predominantly motivated by moral inclinations to act in a just manner. Other theories suggest that individuals want to maintain a positive self-image and weigh up the costs of losing that self-image against the benefits of cheating (e.g., Ariely et al. (2009); Fischbacher and Föllmi-Heusi (2013)). Next to these theories a plethora of experiment on cheating behaviour shows that it is influenced by various contextual and situational factors (e.g., Schild et al. (2019)), including monetary incentives (e.g., Gneezy et al. (2011); Conrads et al. (2013)), time constraints (e.g., Van der Cruyssen et al. (2020); Gunia et al. (2012); Shalvi et al. (2012)).

While external factors have been the primary focus of research in economics concerning the mitigation of cheating behaviour, the role of subtle environmental cues in influencing such behaviour has received relatively less attention in the literature. Understanding the influence of more subtle environmental cues is of importance for two main reasons. First, it is essential to comprehend the efficacy of moral awareness cues and people's behavioural tendencies when faced with the temptation to deceive. Moral awareness, defined as a precondition of moral action, can aid in the promotion of ethical behaviour. In particular, the implementation of systematic, context-specific moral awareness cues can strengthen the reinforcement of ethical conduct. Second, in many contexts, designers of decision-making environments have only the option of subtle cues, since the environment does not allow for the implementation of monetary incentives or the variation of other situational factors. This is particularly true for websites and online environments where brief moral awareness prompts can be a means of encouraging ethical behaviour, e.g. to report honestly or to donate. In doing so, we follow a call for research by Grym et al. (2016), who argue that it would be beneficial to understand the effects of different moral reminders in different testing situations, such as online and computer-based.

In this paper, we test the influence of moral awareness cues that refer to the concept of moral awareness, which originates from Rest's four - component model that describes decision-making and moral action as a gradual process (Rest, 1986). Moral awareness, defined as a precondition of moral action, can aid in the promotion of ethical behaviour. In particular, the implementation of systematic, context-specific moral awareness cues can strengthen the reinforcement of ethical conduct. In addition, Rest (1986) defined moral awareness as an interpretive process through which an individual recognizes that a moral problem exists in a situation and that a moral standard or principle is relevant to a set of circumstances. Moral awareness refers to an individual's ability to perceive the moral dimensions of a situation that may be ethically ambiguous. This idea is based on Rest's model of ethical decision-making and moral action, which suggests that the process involves several components, including moral awareness. The model proposes that moral awareness is the first step in the process and involves recognising the moral nature of a situation. Following moral awareness, individuals make moral judgments, determine their moral intentions and take moral action. According to Jones (1991) and Rest (1986), moral awareness is a fundamental aspect of ethical behaviour and serves as a critical foundation for the subsequent stages of ethical decision making and moral action. Rest (1986) specifies certain prerequisites that are necessary for an individual to possess moral awareness. These consist of assessing the different possible actions that could be taken in a specific situation, acknowledging the individuals who may be influenced by each alternative, and envisioning how these individuals might react to different options (Rest, 1986). As a result, Rest contends that moral awareness is contingent on three critical elements: (i) understanding the ethical ramifications of different actions, (ii) identifying the people who could be impacted, and (iii) anticipating their responses.

In our well-powered online experiments on MTurk (n=551), participants had to flip a coin and report the result (Dickinson and McEvoy, 2021). The incentives were such that reporting heads would increase their chances of winning a monetary reward (Bucciol and Piovesan, 2011; Fischbacher and Föllmi-Heusi, 2013) and decrease a donation that is made on behalf of the participants to UNICEF. We design treatments based on the three theorized elements of moral awareness (Rest, 1986) and compare the frequency of reported heads with a control treatment. While the control treatment contained no cues, the other treatments contained one or more of the following: a specific moral statement designed to promote honest behaviour, reminders of how participants' decisions impacted on relevant parties, information about the reactions of affected parties, or a monitoring cue designed to increase moral awareness. Following the coin flip task, we conducted a survey to collect information on participants' socio-economic characteristics and Big Five traits (Conrads et al., 2013).

The results show that none of the moral awareness cues had any effect on participants' cheating behaviour. While there is evidence that participants cheated in all treatments, the frequency of successfully reported coin tosses (heads) did not differ between treatments. Exploratory analyses show that those participants who had previously donated to UNICEF cheated more than those who had no previous interaction with this charity. Exploratory analyses also show that older and more agreeable participants tended to report more honestly, while more extroverted participants cheated slightly more.

Our experimental setup aligns with some more subtle approaches in the literature to increase honest behaviour that can broadly be classified into three concepts: honesty oaths/commitment, honesty nudges, and moral reminder.

Honesty oaths and commitments require participants to make explicit promises to be truthful and honest. Often this is accompanied with a signature or some form of commitment. Honesty oaths are designed to make individuals feel more accountable for their actions and encourage them to act in a way that aligns with their stated commitment to honesty. Kristal et al. (2020); Brink et al. (2019) found that having participants sign a honesty oath before a cheating task did not reduce cheating compared to signing the oath after the task. However, other studies, such as Beck et al. (2020), have shown the effectiveness of honesty oaths in reducing dishonest behaviour, such as lying in a dice game, after signing the oath.

Honesty nudging through norm salience appears to be a more recent approach to foster honest behaviour. Based on experimental data gathered from an online platform, Dimant et al. (2020) discovered that normative messages, which convey the actions most individuals endorse in similar situations, do not decrease dishonesty. Similarly, in a field study Brudermann et al. (2015) did not find any effect of social norms on cheating. In their field setting, an honour system for the sale of newspapers displaying that a "majority of the readers pay for their copy" before the purchasing decision did not lead to less cheating.

Closest to our experiment is the literature on moral reminders[1]. Moral reminder usually involves the use of language or images that activate moral values or concepts in the mind of the individual. For example, reminding someone of their duty to behave honestly in a specific situation should encourage honest behaviour. Examples for the effectiveness of moral reminders have been the work by laboratory studies of Mazar et al. (2008) who study for example the reading the ten commandments before a cheating task makes people report more honestly. On the contrary, Schild et al. (2019) find that moral reminders do not decrease dishonesty. In

[1]See Hertwig and Mazar (2022) for a more broad taxonomy of honesty interventions.

their online experiment, participants needed to unscramble a sentence which contained ethical primes before engaging in a cheating task.

To sum up, the strand of literature shows a mixture of findings regarding the effectiveness of moral awareness on honesty. Our study contributes to the literature on moral awareness by utilising Rest's (1986) theory to identify the key antecedents of moral awareness and designing treatments based on them. By building on an established theory and leveraging these antecedents, our experiment should shed light on the factors that impact moral awareness and provide insights into how individuals decide more honestly.

The remainder of this paper is organised as follows. In the following section, we summarise the experimental procedures and treatments. In Section 3, we present our main results and regressions. In Section 4, we present the conclusion.

2 Experiment

2.1 Coin-flip task

To examine cheating behaviour in our experiment, we employ the coin-flip paradigm. The coin-flip paradigm, devised by (Bucciol and Piovesan, 2011) was used in our experiments. In which, participants were asked to flip a coin, observe the outcome, and report. Participants' payoffs depended on their reports. Regarding monetary incentives, participants were informed that they would be awarded a \$50 lottery ticket for each reported heads outcome, and that UNICEF would be awarded a \$50 lottery ticket for each reported tails outcome. It is explained on the decision screen that the donations will be awarded to a program fighting malnutrition of children and that the award could result in saving 100 lives of malnourished children (Appendix – Figure A.2). They were assigned the task of physically flipping a coin of their choice 20 times and reporting the outcome of each flip (i.e., heads or tails) in the digital questionnaire provided to them via the website. The participants flipped the coin in a context in which the outcomes, which were random, were observable only by them. The participants then reported the outcomes to the experimenter responsible for paying the monetary incentive based on the reported numbers.

To illustrate the strategic situation, let us consider the following case, where all participants in a treatment report honestly and a treatment has 100 participants. Thus, on average, a participant creates 10 tickets for the lottery. The probability of a ticket being selected for a given participant is 1/1000, so the expected value is $\$50 \times \frac{1}{1000} = \$0.05$. It is easy to see that there is

an incentive for a participant to report heads and to misreport the observed outcome. Suppose that one participant reports heads in each round. This increases the number of tickets by 10 to 10010. The expected value is $\$50 \times \frac{20}{10010} = \$0.09945$.

In our study, we made a conscious decision not to disclose the specific odds, such as 1/1000, to the participants. Our primary motivation for this choice was to simulate real-world scenarios where individuals frequently make decisions without having complete information or without being explicitly aware of the exact probabilities. We believe that by mirroring such real-world uncertainties, our experiment provides insights that might be more generalizable to everyday decision-making situations people encounter.

2.2 Treatments

Our goal was to design treatments that activate one of the theorized prerequisites of moral awareness. Moral awareness is defined by an individual's sensitivity to recognising the moral nature of an ethically ambiguous situation. The concept of moral awareness derives from Rest's four-component model, which describes ethical decision making and moral action as a gradual process (Rest, 1986). This process includes moral awareness, which is recognising the moral nature of a situation; moral judgement, which is deciding what is morally right in the situation; moral intention, which is deciding to prioritise moral values over other values; and moral action, which is following through on moral intention with moral behaviour (Jones, 1991; Rest, 1986). Within this definition, moral awareness is described as the critical foundation for ethical behaviour.

As prerequisites for moral awareness, Rest mentions that the person must have interpreted different possible actions within the specific situation, identified persons affected by different courses of action, and anticipated corresponding reactions of affected parties to different courses of action (Rest, 1986). Thus, according to Rest, moral awareness has three prerequisites: (i) interpretation of possible actions and inherent morality of the decision, (ii) identification of affected parties, and (iii) anticipation of affected parties' reactions.

To test whether one of these preconditions alone is capable of raising moral awareness and leading to moral action, we design three different treatments. Treatment 1 is the least vivid and most abstract cue and refers to the first prerequisite for moral awareness, i.e. becoming aware of the inherent morality of the decision that the participant is about to make. Before starting the task, participants read *"Be aware of the moral aspects inherent in your decision to report heads or tails after each coin flip!"* We chose this wording to increase awareness of the moral

aspects of the coin flip task, deliberately avoiding the use of the word honesty or truthfulness so as not to steer participants in a particular direction, but to make it clear that they were facing a moral dilemma. Treatment 2 has a cue of a medium level of vividness and a medium level of abstraction and relates to the second prerequisite for moral awareness, i.e. the recognition that one's decisions affect other parties. Before starting the task, participants read *"Imagine you win $50, and imagine 100 suffering children that can be fed with $50!"*. We chose this wording to make participants aware of the consequences of their decisions for the people affected. Treatment 3's cue has a high level of vividness and a low level of abstraction and refers to the third precondition of moral awareness, i.e. increasing the anticipation of the reaction of the persons concerned. Before starting the task, participants read *"Imagine the joy and gratitude of 100 smiling children receiving food they desperately need!"* We chose this wording to make it clear to the participants that their decisions will have consequences for the people they affect.

Independently of the moral awareness theory, we conducted a fourth treatment (T4) in which we attempted to manipulate participants' perception of being observed. Before the task, participants read *"Envision 100 children who could benefit, watching you while you are flipping the coin and reporting the outcomes."* Typically, increased perceived monitoring can lead to more ethical behaviour (Bateson et al., 2006). We added this treatment and chose this wording to have another benchmark treatment that included a manipulation of perceived monitoring but was similar in wording and content to the other treatments.

Building on this theory, we hypothesised that moral awareness cues targeting the third level (T3) would be the most effective in reducing cheating behaviour, while cues targeting the first level (T1) would be the least effective, and cues targeting the second level (T2) would have a moderate effect.

In our study, we aimed to create a scenario that would resemble real-life situations, particularly online environments such as the website of UNICEF. To achieve this, we used phrasing that is commonly found in such contexts, such as "desperate need of" and "suffering children." Our goal was to create a realistic and relatable scenario that would increase the ecological validity of our study and make it more applicable to real-world situations. By using wording that is commonly found in practical contexts, we aimed to create a situation that would be more engaging and meaningful for our participants, and would therefore increase the validity and generalizability of our findings.

The introduction of the UNICEF donation was not merely to add a charitable dimension. Instead, it was strategically chosen to underscore the broader societal implications of cheating,

emphasising the negative consequences of one's decisions on others. This approach was rooted in our aim to highlight that unethical behaviour often has repercussions that extend beyond the immediate context, affecting vulnerable individuals or groups in society. By making participants aware of the tangible impact of their honesty (or lack thereof) on a charitable cause, we aimed to amplify the moral dimension of their choices, using the charitable outcome as a representation of the broader societal consequences of individual actions.

2.3 Power analysis and sample size

To determine our sample size we used the software GPower. To observe a medium sized effect (d= 0.500) with a power of 95% (α = 0.050, one-tailed), the 'A priori power analysis' yielded a target sample size of at least N=88 per treatment. Since we were not sure how many participants would drop out during the study and we wanted to avoid starting the study multiple times, we waited for at least 100 participants per treatment to have started the survey and then decided to close the study. We ended up having 551 observations for the entire study. The details of our variables statistics and randomization test can be found at Appendix A.1.

2.4 Procedures

After the first ten coin flips using their Amazon MTurk ID and MTurk Prime ID, participants were asked to provide their socioeconomic information, such as age and gender. The study also collected data on the results of personality testing, including the Big Five personality trait assessment (Psychometrics, 2018), a domain-specific risk-taking (DOSPERT) assessment (Blais and Weber, 2006; Weber et al., 2002), and a willpower assessment (Change, 2018). We also asked participants if they had donated to UNICEF in the past to control for their preference for the charity. After providing the information, the participants performed and reported the results of the remaining ten coin flips. Eventually, all the relevant information on the participants' performance and characteristics were collected for further analysis. The study participants were randomly assigned to one of four treatments or the baseline[2].

3 Results

Before testing our hypotheses and to examine whether the participants lied across the treatments, we tested whether the actual number of reported heads outcomes per treatment was significantly higher than the expected number (i.e., 10). The results of the t-tests indicate that

[2]We acknowledged that variances in subject pools can exist between online platforms and traditional lab studies. However, our approach aligns with findings from Snowberg and Yariv (2021), indicating behavioural comparability between online and in-person experiments.

in all of the treatments, the participants cheated and significantly inflated their reports ($p < 0.050$ for all treatments and baseline). The results further reveal that in all of the treatments, including in the baseline, the proportion of participants who reported an above-average number of heads outcomes was higher than the proportion of participants who were 'potentially honest' (reporting less than 10 heads) ($p < 0.001$, tests of proportions).

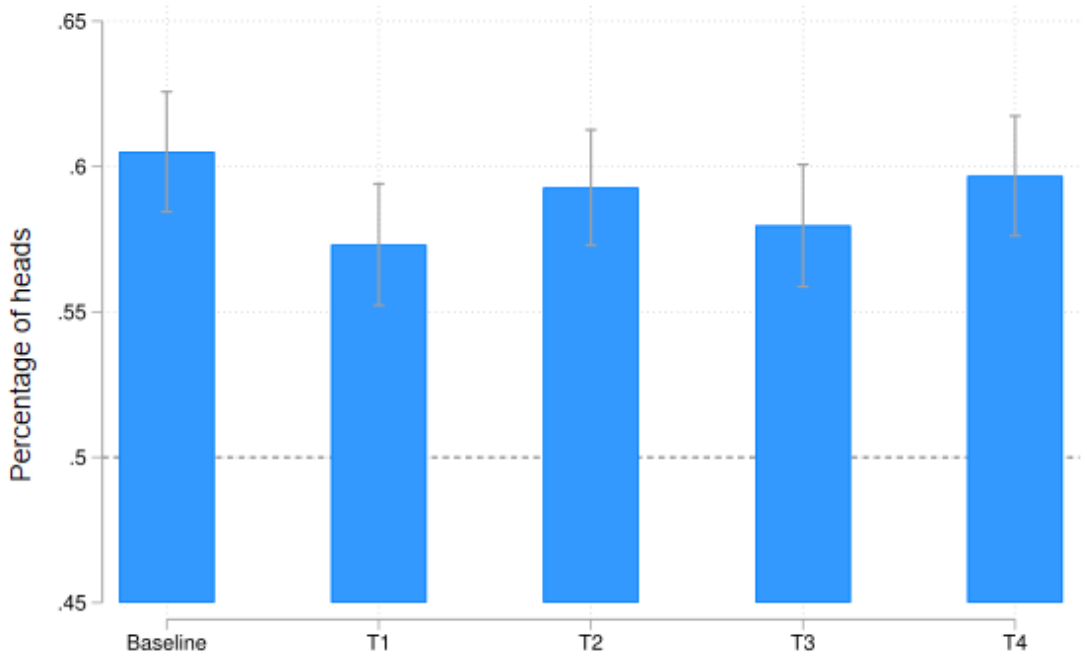
3.1 Hypotheses test

Figure 1 shows the share of heads reports (panel (a)) and the probability density function of the number of heads reported (panel (b)) for all treatments. Pairwise comparisons of the number of reported heads between baseline and each treatment show no significant difference (Baseline and T1, $p = 0.179$; Baseline and T2, $p = 0.610$; Baseline and T3, $p = 0.306$; Baseline and T4, $p = 0.681$, pairwise comparisons of means). Again, when comparing T1 with T2 ($p = 0.388$), T2 with T3 ($p = 0.590$) and T3 with T4 ($p = 0.536$), there is no statistically significant difference in the number of heads reported, which is done with pairwise comparisons of means. The probability density function appears to be nearly identical between the baseline and all treatment groups, with minor variations observed in the intervals between 7 and 10 reported heads. This indicates that there may be a subtle shift in distribution among the participants. To sum up, it appears that none of our moral awareness cues was effective in leading to less cheating compared to a situation with no cue.

In addition, the fact that the treatment was only implemented before the first round of the task, rather than before each round, could be another potential explanation for our null results. To test this possibility, we analysed the share of head reports in each treatment compared to the baseline and found no significant effects by using the regression ($p > 0.356$, mean t-test). We also tested each pairwise t-test between each treatment and the baseline (Baseline - T1, $p > 0.163$; Baseline - T2, $p > 0.621$; Baseline - T3, $p > 0.287$; Baseline - T4, $p > 0.657$, mean t-test). This suggests that participants did not change their behaviour compared to the baseline despite being given an immediate moral awareness cue in the first round. This finding raises questions about the longevity of the effect of our treatment and the need for ongoing reminders to maintain moral awareness and encourage honest behaviour.

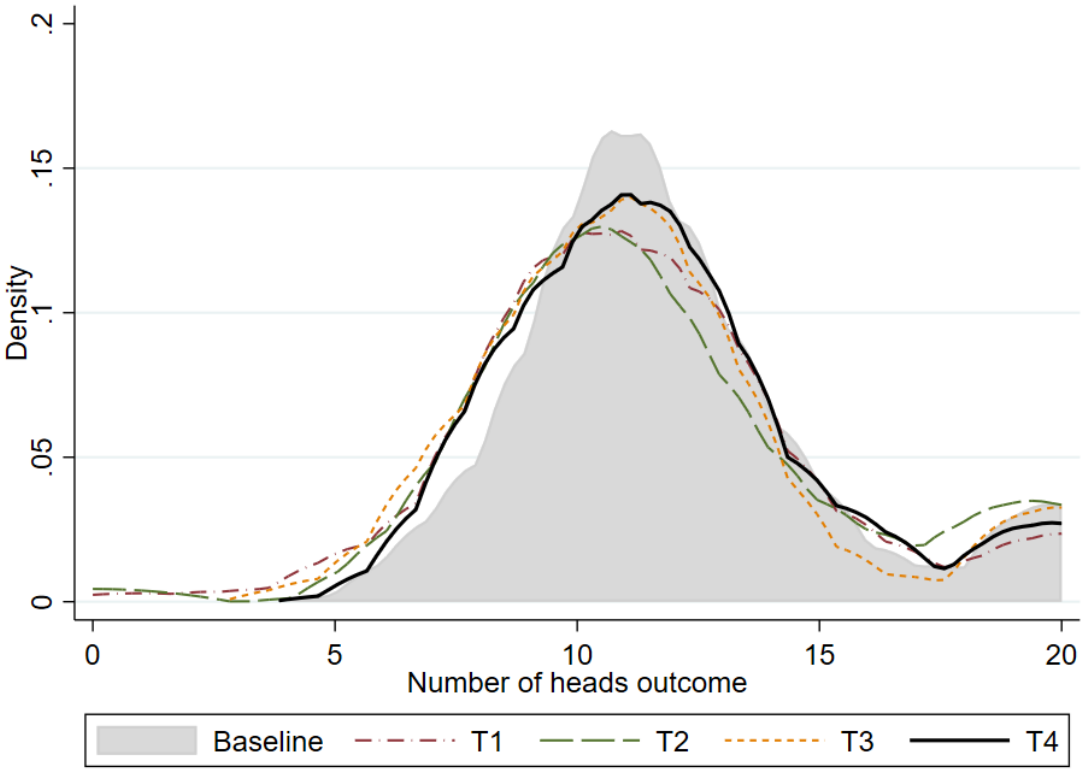
3.2 Supplemental analyses

To validate our findings, we controlled for the demographic factors and personality traits using different regressions. Table 1 presents the regression results on the determinants across the treatments, including the socioeconomic factors and personality traits associated with the



(a) Percentage of heads in each treatment

Error bars indicate 95% confidence interval of the mean.



(b) Probability Density Function

Figure 1: Head reports over five treatments

numbers of 'heads' outcomes reported. First, in models (1) and (2) we predict the number of heads reported. In model (1), in line with our pairwise comparisons from above, we do not find a significant decrease in the number of heads reported in our treatments, as the insignificant coefficients suggest. The addition of an additional control variable in model (2) does not change this picture either.

In models (3) and (4) we predict the probability of reporting the number of heads as less than or equal to 10. This could be interpreted as an indication that an individual is reporting honestly. As suggested by the treatment dummy for T1 in model (3), we observe some mild evidence that the likelihood of reporting the number of heads equal to or less than ten is slightly higher compared to the baseline. All other treatment dummies remain insignificant. Adding additional controls in model (4) does not change this observation.

In models (5) and (6) we predict the probability of reporting 20, which could be interpreted as an indication that a participant is dishonest. Again, the insignificant treatment dummies suggest that there is no difference in reporting behaviour compared to baseline. The inclusion of additional controls does not change this pattern.

In addition to these comparisons, the regression analyses reveal some interesting findings of an exploratory nature. For example, the positive and significant UNICEF dummy (Model 2 and Model 6) suggests that those participants who donated to UNICEF prior to the experiment were more likely to report heads (UNICEF: n = 30, mean = 13.866) than those participants who did not donate prior to the experiment (n = 521, mean = 11.660, $p = 0.001 < 0.01$, T-test). An untabulated subsample analysis also shows that the null effect of our moral awareness cues is not driven by these participants, as all treatment dummies remain insignificant. As the number of observations of participants who donated to UNICEF before each treatment is rather small (Baseline = 6; T1 = 4; T2 = 11; T3 = 6; T4 = 3), we are cautious in interpreting the result.

In addition, age was found to be the most robust predictor (models 2 and 4), consistent with previous literature (Gerlach et al., 2019). Older participants were more likely to behave truthfully, in line with previous studies (Conrads et al., 2013; Friesen and Gangadharan, 2013; Glätzle-Rützler and Lergetporer, 2015). Two theories may explain this phenomenon: theories of risk-taking behaviour (Josef et al., 2016; Mata et al., 2016) and the effects of maturity on personality (Roberts et al., 2006).

Finally, we used personality traits (Psychometrics, 2018; Rammstedt and John, 2007) to gain a better understanding of how their personalities influenced cheating behaviours. We found the

factor of agreeableness had a negative relationship with the number of heads reported (models (2) and (6)), while extraversion had a positive relationship with the number of heads outcomes reported (model (2)). This implies that participants who were typically polite and likable (as indicated by having a high agreeableness score) tended to cheat less, in line with the findings of previous theoretical research (Graziano and Tobin, 2019; Nathanson et al., 2006). The positive association of extraversion with cheating behaviour is in line with the findings of Cizek (1999). In our experiment, we found that participants who were more outgoing and inclined towards sensation-seeking were more likely to report a high frequency of heads outcomes. However, it's important to note that these participants didn't report the maximum number of heads, as the coefficient in Model (6) turned out to be insignificant. Interestingly, our findings diverged from previous experiments that used the Big Five personality traits, as we did not find any predictive power of conscientiousness, neuroticism, and openness to experience for cheating behaviour in our sample. This suggests that factors other than these traits may be more relevant in influencing cheating behaviour (Feldman, 1964; Hogan and Hogan, 1989).

Table 1: Regressions of reported outcomes from the experiment

Variables	Head outcomes reported		Likelihood of reporting ≤ 10		Likelihood of reporting = 20	
	Model (1)	**Model (2)**	**Model (3)**	**Model (4)**	**Model (5)**	**Model (6)**
T1	-0.682	-0.552	0.775*	0.756*	-0.138	0.003
	[-1.270]	[-1.072]	[2.334]	[2.228]	[-0.272]	[0.005]
T2	-0.240	-0.304	0.578	0.586	0.219	0.230
	[-0.456]	[-0.597]	[1.736]	[1.694]	[0.474]	[0.480]
T3	-0.496	-0.360	0.585	0.533	0.136	0.250
	[-0.918]	[-0.693]	[1.722]	[1.522]	[0.283]	[0.485]
T4	-0.206	-0.193	0.535	0.557	-0.020	-0.014
	[-0.385]	[-0.373]	[1.579]	[1.616]	[-0.041]	[-0.027]
1 if UNICEF		1.963**		-0.075		1.109**
		[2.687]		[-0.151]		[2.327]
Age		-0.057***		0.027**		-0.036
		[-3.939]		[3.116]		[-1.930]
1 if female		-0.462		0.195		-0.125
		[-1.311]		[0.873]		[-0.376]
Extraversion		0.053**		-0.021		0.032
		[2.606]		[-1.665]		[1.410]
Agreeableness		-0.065*		0.023		-0.053*
		[-2.399]		[1.339]		[-1.972]
Conscientiousness		0.012		-0.011		0.010
		[0.445]		[-0.619]		[0.367]
Neuroticism		-0.016		0.008		-0.024
		[-0.765]		[0.559]		[-1.293]
Openness to Experience		-0.019		0.022		-0.009
		[-0.735]		[1.338]		[-0.337]
Constant	12.225***	16.079***	-1.609***	-3.576***	-2.398***	0.221
	[32.121]	[15.548]	[-6.228]	[-5.024]	[-6.881]	[0.221]
Observation	551	551	551	551	551	551
Pseudo R2	0.0007	0.02	0.01	0.05	0.002	0.08
Log likelihood	-1469.156	-1443.02	-307.049	-296.013	-162.641	-149.316
F-stat	1.97	54.24	5.7	27.38	0.700	28.33

Notes: The symbols *, **, and *** indicate 5%, 1%, and 0.1% two-tailed significance levels, respectively. Robust standard errors are presented in parentheses. The reference group is the baseline treatment. Tobit regressions were employed for models (1) and (2) with a lower limit of (0) and an upper limit of (20). Logit regressions were used for the remaining models. UNICEF is a dummy variable indicating if the participant had previously donated to UNICEF, 0 otherwise. The personality traits are extraversion, agreeableness, conscientiousness, neuroticism, and openness to experience (Psychometrics, 2018; Rammstedt and John, 2007). The description of these factors can be found in Rammstedt and John (2007). F-stats represent the hypothesis with multiple coefficients simultaneously being 0.

3.3 Discussion

There are at least four possible explanations for the null result of the moral awareness cues. First, the context of the online experiment did not fully raise participants' moral standards through the interpretive process, as Rest's theory suggests. Although Rest's theory explains the necessary preconditions for moral awareness, it is silent on how these preconditions can be activated. Our - admittedly rather mild - one-sentence language intervention might activate moral awareness in some individuals, but not enough to change the average behaviour. In this context, the incentives to misreport the outcome of the coin flip were rather low and might limit our ability to detect a large effect. Although our study was well powered, the benefits of cheating in our study were rather small compared to existing studies.

Second, it may be that participants did not perceive the decision to misreport the coin flip and the resulting consequence for UNICEF as unethical as we thought ex-ante. It may well be that the distance between the third party involved, UNICEF, and the participant was too great and therefore the negative consequence for UNICEF was not perceived as unethical, which in turn did not lead to more honest reporting. Some evidence for this comes from the fact that those participants who had previously donated to UNICEF cheated even more, which may indicate that even those participants with a stronger connection to UNICEF did not care about the consequences of misbehaviour for UNICEF's activities in our experiment.

Third, while mTurk participants are a convenient source of research participants, they may not be representative of the general population. For example, research has shown that mTurk participants tend to be younger, more educated and more politically liberal than the general population. In addition, mTurk participants are often regular users of online platforms, which may influence their behaviour and responses (Chandler et al., 2019) Given these factors, it is possible that the participants in our study may not be representative of the general U.S. population and that the mild interventions used in the study may not be effective for everyone. Therefore, the results of the study may not be generalisable to other populations or contexts.

Fourth, the coin flip paradigm we used to measure cheating may not have been sensitive enough to detect small changes in cheating, especially if the moral awareness cues we used were relatively weak. The coin flip paradigm has several limitations. For example, the potential gains from cheating in this paradigm may not be large enough to motivate participants to cheat, especially if they perceive that the risk of being caught is high. In addition, the coin flip paradigm may not fully capture the complexity of moral decision-making, which may involve multiple factors beyond simple reward incentives. As a result, it is possible that the coin-flip paradigm

was not sensitive enough to detect small changes in cheating behaviour that may have been induced by the moral awareness cues we used.

4 Conclusion

We tested Rest's (1986) theory that individuals are more likely to behave ethically when they are aware of moral principles and standards. We were interested in whether subtle reminders of moral awareness, based on this theory, would reduce the likelihood of cheating behaviour. Using an online experiment with 551 participants, we found that cues to moral awareness did not reduce the likelihood of cheating. Although we did not find a statistically significant effect of any of these cues in our context, this does not mean that moral cues in general have no effect on cheating behaviour. Further research is needed to explore these possibilities and to better understand the relationship between moral cues and cheating behaviour.

Still, our findings have implications for decision making in online environments. One practical implication of our null result is that simply reminding individuals of moral principles or standards may not be sufficient to prevent unethical behaviour. Either increasing the frequency of reminders or other interventions, such as increasing the perceived risk of being caught or promoting a culture of honesty and integrity, may be necessary to effectively deter cheating behaviour. Future research should consider repeating these statements to participants in each coin flip trial.

References

Abeler, J., Nosenzo, D., Raymond, C., 2019. Preferences for truth-telling. Econometrica 87, 1115–1153.

Ariely, D., Bracha, A., Meier, S., 2009. Doing good or doing well? Image motivation and monetary incentives in behaving prosocially. American Economic Review 99, 544–55.

Bateson, M., Nettle, D., Roberts, G., 2006. Cues of being watched enhance cooperation in a real-world setting. Biology Letters 2, 412–414.

Beck, T., Bühren, C., Frank, B., Khachatryan, E., 2020. Can honesty oaths, peer interaction, or monitoring mitigate lying? Journal of Business Ethics 163, 467–484.

Blais, A.R., Weber, E.U., 2006. A domain-specific risk-taking (dospert) scale for adult populations. Judgment and Decision Making 1, 33–47.

Brink, W.D., Eaton, T.V., Grenier, J.H., Reffett, A., 2019. Deterring unethical behavior in online labor markets. Journal of Business Ethics 156, 71–88.

Brudermann, T., Bartel, G., Fenzl, T., Seebauer, S., 2015. Eyes on social norms: A field study on an honor system for newspaper sale. Theory and Decision 79, 285–306.

Bucciol, A., Piovesan, M., 2011. Luck or cheating? A field experiment on honesty with children. Journal of Economic Psychology 32, 73–78.

Capraro, V., Halpern, J.Y., Perc, M., 2022. From outcome-based to language-based preferences. Journal of Economic Literature .

Capraro, V., Perc, M., 2021. Mathematical foundations of moral preferences. Journal of the Royal Society Interface 18, 20200880.

Chandler, J., Rosenzweig, C., Moss, A.J., Robinson, J., Litman, L., 2019. Online panels in social science research: Expanding sampling methods beyond mechanical turk. Behavior research methods 51, 2022–2038.

Change, T.W.T., 2018. Willpower assessment. Willpower Assessment website URL: http://www.thewilltochange.com/the-willpower-program/willpower-assessment/.

Cizek, G.J., 1999. Cheating on tests: How to do it, detect it, and prevent it. Routledge.

Conrads, J., Irlenbusch, B., Rilke, R.M., Walkowitz, G., 2013. Lying and team incentives. Journal of Economic Psychology 34, 1–7.

Van der Cruyssen, I., D'hondt, J., Meijer, E., Verschuere, B., 2020. Does honesty require time? Two preregistered direct replications of experiment two of Shalvi, Eldar, and Bereby-Meyer (2012). Psychological Science 31, 460–467.

Dickinson, D.L., McEvoy, D.M., 2021. Further from the truth: The impact of moving from in-person to online settings on dishonest behavior. Journal of Behavioral and Experimental Economics 90, 101649.

Dimant, E., Van Kleef, G.A., Shalvi, S., 2020. Requiem for a nudge: Framing effects in nudging honesty. Journal of Economic Behavior & Organization 172, 247–266.

Feldman, S.E., 1964. College cheating as a function of subject and situational variables. Journal of Educational Psychology 55, 212–218.

Fischbacher, U., Föllmi-Heusi, F., 2013. Lies in disguise — An experimental study on cheating. Journal of the European Economic Association 11, 525–547.

Friesen, L., Gangadharan, L., 2013. Designing self-reporting regimes to encourage truth telling: An experimental study. Journal of Economic Behavior & Organization 94, 90–102.

Gerlach, P., Teodorescu, K., Hertwig, R., 2019. The truth about lies: A meta-analysis on dishonest behavior. Psychological Bulletin 145, 1–44.

Glätzle-Rützler, D., Lergetporer, P., 2015. Lying and age: An experimental study. Journal of Economic Psychology 46, 12–25.

Gneezy, U., Meier, S., Rey-Biel, P., 2011. When and why incentives (don't) work to modify behavior. Journal of economic perspectives 25, 191–210.

Graziano, W.G., Tobin, R.M., 2019. Theoretical conceptualizations of agreeableness and antagonism. The Handbook of Antagonism 1, 127–139.

Grym, J., Liljander, V., et al., 2016. To cheat or not to cheat? The effect of a moral reminder on cheating. Nordic Journal of Business 65, 18–37.

Gunia, B.C., Wang, L., Huang, L., Wang, J., Murnighan, J.K., 2012. Contemplation and conversation: Subtle influences on moral decision making. Academy of Management Journal 55, 13–33.

Hertwig, R., Mazar, N., 2022. Toward a taxonomy and review of honesty interventions. Current Opinion in Psychology , 101410.

Hogan, J., Hogan, R., 1989. How to measure employee reliability. Journal of Applied Psychology 74, 273–279.

Jones, T.M., 1991. Ethical decision making by individuals in organizations: An issue-contingent model. Academy of Management Review 16, 366–395.

Josef, A.K., Richter, D., Samanez-Larkin, G.R., Wagner, G.G., Hertwig, R., Mata, R., 2016. Stability and change in risk-taking propensity across the adult life span. Journal of Personality and Social Psychology 111, 430–450.

Kristal, A.S., Whillans, A.V., Bazerman, M.H., Gino, F., Shu, L.L., Mazar, N., Ariely, D., 2020. Signing at the beginning versus at the end does not decrease dishonesty. Proceedings of the National Academy of Sciences 117, 7103–7107.

Leib, M., Köbis, N., Soraperra, I., Weisel, O., Shalvi, S., 2021. Collaborative dishonesty: A meta-analytic review. Psychological Bulletin 147, 1241.

Mata, R., Josef, A.K., Hertwig, R., 2016. Propensity for risk taking across the life span and around the globe. Psychological Science 27, 231–243.

Mazar, N., Amir, O., Ariely, D., 2008. The dishonesty of honest people: A theory of self-concept maintenance. Journal of Marketing Research 45, 633–644.

Nathanson, C., Paulhus, D.L., Williams, K.M., 2006. Predictors of a behavioral measure of scholastic cheating: Personality and competence but not demographics. Contemporary Educational Psychology 31, 97–122.

Psychometrics, O., 2018. The big five personality test. Open Psychometrics website URL: https://openpsychometrics.org/printable/big-five-personality-test.pdf.

Rammstedt, B., John, O.P., 2007. Measuring personality in one minute or less: A 10-item short version of the big five inventory in english and german. Journal of Research in Personality 41, 203–212.

Rest, J.R., 1986. Moral development: Advances in research and theory. Praeger Publisher.

Roberts, B.W., Walton, K.E., Viechtbauer, W., 2006. Patterns of mean-level change in personality traits across the life course: A meta-analysis of longitudinal studies. Psychological Bulletin 132, 1–25.

Schild, C., Heck, D.W., Ścigała, K.A., Zettler, I., 2019. Revisiting REVISE: (Re)testing unique and combined effects of REminding, VIsibility, and SElf-engagement manipulations on cheating behavior. Journal of Economic Psychology 75, 1–14.

Shalvi, S., Eldar, O., Bereby-Meyer, Y., 2012. Honesty requires time (and lack of justifications). Psychological Science 23, 1264–1270.

Snowberg, E., Yariv, L., 2021. Testing the waters: Behavior across participant pools. American Economic Review 111, 687–719.

Tappin, B.M., Capraro, V., 2018. Doing good vs. avoiding bad in prosocial choice: A refined test and extension of the morality preference hypothesis. Journal of Experimental Social Psychology 79, 64–70.

Weber, E.U., Blais, A.R., Betz, N.E., 2002. A domain-specific risk-attitude scale: Measuring risk perceptions and risk behaviors. Journal of Behavioral Decision Making 15, 263–290.

Figure Appendix

Figure A.1: Introductory page (Experiment documentation)

Notes: This picture shows the first page of the experiment.

Figure A.2: The example of moral awareness cues

Notes: This picture resembles to Treatment 4, implying to use a monitoring effect.

Table A.1: Statistics summary and randomization test between control and treated group

	(1)	(2)	(3)	(4)	(5)	(2)–(1)	(3)–(1)	(4)–(1)	(5)–(1)
	Baseline	**T1**	**T2**	**T3**	**T4**		**Pairwise t-test**		
Variable	Mean/(SE)	Mean/(SE)	Mean/(SE)	Mean/(SE)	Mean/(SE)	Mean diff	Mean diff	Mean diff	Mean diff
Heads reported	12.102	11.440	11.856	11.594	11.900	-0.661	-0.246	-0.508	-0.202
	(0.320)	(0.347)	(0.374)	(0.353)	(0.322)				
UNICEF	0.056	0.037	0.093	0.057	0.027	-0.019	0.038	0.001	-0.028
	(0.022)	(0.018)	(0.027)	(0.023)	(0.016)				
Age	38.954	38.780	40.254	41.028	38.036	-0.174	1.301	2.075	-0.917
	(1.164)	(1.126)	(1.139)	(1.148)	(1.117)				
Gender	0.537	0.550	0.508	0.538	0.664	0.013	-0.029	0.001	0.127*
	(0.048)	(0.048)	(0.046)	(0.049)	(0.045)				
Extraversion	16.389	17.706	18.958	17.311	17.118	1.318	2.569**	0.922	0.729
	(0.857)	(0.845)	(0.905)	(0.852)	(0.872)				
Agreeableness	28.222	29.881	28.390	28.708	27.564	1.659	0.168	0.485	-0.659
	(0.766)	(0.657)	(0.658)	(0.649)	(0.648)				
Conscientiousness	28.926	29.257	28.458	27.632	27.764	0.331	-0.468	-1.294	-1.162
	(0.652)	(0.664)	(0.621)	(0.708)	(0.681)				
Neuroticism	23.028	25.459	26.339	23.981	24.255	2.431*	3.311***	0.953	1.227
	(0.958)	(0.917)	(0.809)	(0.998)	(0.932)				
Openness	27.222	28.110	27.322	27.075	27.000	0.888	0.100	-0.147	-0.222
	(0.720)	(0.622)	(0.554)	(0.703)	(0.655)				
N	108	109	118	106	110	217	226	214	218

Notes: The symbols *, **, and *** indicate 5%, 1%, and 0.1% two-tailed significance levels, respectively. The mean difference shows the t-statistics test of the mean difference between the control and treated groups. The majority of our participants are from the United States (96%) and other countries (such as Australia, Canada, China, Russia, Serbia, Vietnam, etc.).