Zoe Rahwan, Oliver P. Hauser, Ewa Kochanowska and Barbara Fasolo

High stakes: a little more cheating, a lot less charity

Article (Accepted version) (Refereed)

Original citation:
Rahwan, Zoe and Hauser, Oliver P. and Kochanowska, Ewa and Fasolo, Barbara (2018) *High stakes: a little more cheating, a lot less charity*. Journal of Economic Behavior and Organization. ISSN 0167-2681 (In Press)
DOI: 10.1016/j.jebo.2018.04.021

High stakes: a little more cheating, a lot less charity

Zoe Rahwan[1,2], Oliver P. Hauser[2,3,4], Ewa Kochanowska[5], Barbara Fasolo[1] *

Abstract. We explore the downstream consequences of cheating—and resisting the temptation to cheat—at high stakes on pro-social behaviour and self-perceptions. In a large online sample, we replicate the seminal finding that cheating rates are largely insensitive to stake size, even at a 500-fold increase. We present two new findings. First, resisting the temptation to cheat at high stakes led to negative moral spill-over, triggering a moral license: participants who resisted cheating in the high stakes condition subsequently donated a smaller fraction of their earnings to charity. Second, participants who cheated maximally mispredicted their perceived morality: although such participants thought they were less prone to feeling immoral if they cheated, they ended up feeling more immoral a day after the cheating task than immediately afterwards. We discuss the theoretical implications of our findings on moral balancing and self-deception, and the practical relevance for organisational design.

Keywords: cheating; incentives; moral licensing; moral self-perceptions; pro-social behaviour.

[1] London School of Economics and Political Science, London WC2A 2AE, UK
[2] Harvard Kennedy School, Cambridge, MA 02138, US
[3] Harvard Business School, Boston, MA 02163, US
[4] University of Exeter Business School, Exeter EX4 4PU, UK
[5] IESE Business School, Barcelona 08034, Spain
* Corresponding author: b.fasolo@lse.ac.uk

1. Introduction

Corporate misconduct and unethical behaviour remain a widespread problem in organisations, ranging from large-scale fraud (e.g., Bernie Madoff, Enron) to smaller, everyday unethical behaviours (Zhang et al. 2015, Sezer, Gino & Bazerman 2015). Organisations often rely on compensation schemes to foster better performance among employees but, at the same time, those incentives could also encourage more cheating inadvertently to meet requirements of such schemes. Recently, for example, employees at Wells Fargo who were financially incentivised for every newly opened bank account created over 2 million bank accounts illegally without customers' permission.

Because the size of incentives can affect motives and behaviour in often unexpected ways (Gneezy & List 2014, Gneezy & Rustichini 2000a,b), past research has paid particular attention to the size of incentives and its effects on immediate opportunities for cheating. A somewhat surprising, yet consistent finding is that the size of incentives for behaving unethically does not affect rates of unethical behaviour much in laboratory studies. We take this research in a new direction by looking at the downstream effects of cheating—or, resisting the temptation to cheat—on future moral behaviour. In particular, we explore the behavioural and psychological consequences of providing an immediate opportunity to cheat at varying sizes of stakes on subsequent charitable giving.

Mazar et al. (2008) document that a four-fold increase in incentives did not change average levels of cheating in a laboratory experiment. In a recent large-scale replication, Kajackaite & Gneezy (2017) observe little change in cheating rates as incentives were increased to much higher stakes than previously studied (up to $50 in a U.S. laboratory setting) when there is some chance of cheating being detected.[6] What explains this refusal to cheat more at higher stakes? Mazar et al. (2008) hypothesise that people attempt to maintain a self-concept of an honest person. Thus, they will cheat to increase their payoff but do so only to the extent that it does not negatively affect their moral self-image.

The idea of maintaining a "moral self" has also been proposed in a separate research stream on moral balancing (Miller & Effron 2010; Zhong, Liljenquist & Cain 2009). This research suggests that people balance their good and bad behaviour over time: when people feel they have sufficiently established that they are a moral person, they become more likely to engage in immoral behaviour in the future ("moral

[6] In another condition, Kajackaite & Gneezy (2017) make detection of cheating almost impossible: the combination of no detection and high stakes does lead to an increase in cheating behaviour, suggesting that an important boundary condition to the literature reviewed here is that participants will cheat more when it is completely impossible to detect their cheating behaviour at the individual level. We discuss the implications of these boundary conditions on our results at the end of the paper.

licensing”). For example, Effron et al. (2009) demonstrate that participants who could endorse Barack Obama were more likely to feel licensed in a subsequent decision to choose a white applicant over a black applicant in a hypothetical hiring decision, having previously established their moral credentials as a non-racist moral decision-maker.

Here we combine the literature on stake size on immediate cheating opportunities with research on moral balancing to explore whether different levels of incentives affect subsequent moral behaviour. We hypothesise that, when given the opportunity to cheat, participants will resist this temptation at low and high stakes; however, resisting the temptation to cheat at high stakes will be psychologically taxing and provide participants with a plausible reason to excuse future transgressions of moral behaviour, exhibiting a moral licensing effect.

To explore the psychological costs and benefits of dynamic moral behaviour, we measure participants' self-perceptions of their morality. While cheating and donations are behavioural, incentive-compatible outcomes, self-reports are simply stated views – however, as such they inform how people (would like to) view themselves and provides an insight into the psychological processes of how cheating is dealt with. We predict that cheating a little does not affect moral self-perceptions; only maximal cheating—an unambiguous signal of immoral behaviour—negatively affects self-perceptions, an effect that persists (or even worsens) over time.

To test these predictions, we recruit a large-scale sample of participants (*N*=2,015) on Amazon Mechanical Turk (MTurk). We begin by exploring the role of stake size on dishonest behaviour using a modified version of the “mind game” (Kajackaite & Gneezy 2017, Cohn et al. 2014). We introduce significant variation in incentives, increasing stakes by up to 500 times. The maximum reward in our study is US $50 for our 10 minute online task – a significant amount for workers on MTurk whose median reservation wage has been estimated at US $1.40 per hour (Horton and Chilton, 2010).

In the second part of the study, we explore the downstream effects from honest (or dishonest) behaviour. First, we collect self-perceptions of morality using self-reports. Second, we give participants the opportunity to give to charity, allowing them to choose how much, if any, of their earnings from their first part of the study they wish to donate to a charity of their choice, a measure of whether they engaged in moral licensing by not giving to charity at all or decreasing their charitable donation. Finally, we invited participants back to the study one day later to assess their self-perceptions of their (non-)cheating behaviour the prior day.

We find that participants were not more likely to cheat when the stakes for cheating increased but with rising stakes, they subsequently donated a smaller percentage of their earnings to charity, consistent with a moral balancing account. Moreover, we find that self-perceptions vary over time: while they are stable for honest participants and those who cheat little, we find that maximal cheaters feel less moral one day after the cheating task, independent of the stake size. Surprisingly, however, self-perceptions did not meet expectations: maximal cheaters initially thought they would be less likely to feel guilty after doing wrong, when, in reality, they were the only group of participants that felt worse upon reflecting on their cheating behaviour.

1.1. Dishonesty

Research on dishonesty and unethical behaviour has attracted much attention given its apparent frequency and high costs it imposes for societies. In recent decades, much work has explored the limits of Becker's (1968) utilitarian approach for understanding dishonest behaviour. Under this model, rational individuals weigh the benefits of dishonest behaviour with the chance and consequences of getting caught. Becker's model does not pay explicit regard to psychological costs of dishonest behaviour (e.g. Mazar et al. 2008, Shalvi et al. 2011, Kajackaite and Gneezy, 2017) nor the related social, organisational or political context (e.g. Gneezy 2005, Cohn et al. 2014).

This traditional model of cheating has been challenged by findings that people are generally insensitive to payoffs for dishonesty (e.g. Mazar et al. 2008, Fischbacher & Föllmi-Heusi 2013). Experimental evidence suggests that most people exhibit at least some degree of aversion to lying (Cappelen, Sørensen & Tungodden 2013, Gneezy, Rockenbach & Serra-Garcia 2013, Kajackaite & Gneezy 2015). Abeler et al. (2016) show in a review of 72 cheating studies with a maximum payoff of $50 that individuals are largely insensitive to external stakes for cheating. Moreover, participants displayed almost no increase in cheating, even when stakes rose to as much as $110 (Hilbig and Thielmann 2017). In fact, individuals frequently engage in low-level cheating and commonly resist maximal cheating, leaving on average three-quarters of the maximum possible payoff on the table (Abeler et al, 2016).

Mazar et al (2008) posit a theory of self-concept maintenance to explain why people do cheat but much less than Becker's theory would predict. They argue that individuals attempt to balance the tension between a desire to enrich themselves materially from dishonest behaviour with maintaining a favourable moral self-appraisal. It is therefore possible for individuals to consider themselves honest but nonetheless cheat "only a little bit" because partial dishonesty does not threaten their positive self-image. Gneezy, Kajackaite & Sobel (2017) offer an alternative explanation of the observation that people cheat a little bit: they find that cheating is more common when participants are unobserved rather than observed, and that the frequency of partial lying increases when the maximal outcome is less likely *ex ante*. They argue that social identity – that is, the socially constructed part of an individual's self-concept – shapes what we consider to be appropriate and, consequently, accounts for non-maximal cheating behaviours.

Still, there are occasions where individuals do respond to rewards. Gneezy, Kajackaite and Sobel (2017) and Abeler et al (2016) find more cheating when participants are unobserved rather than observed. Further, many experiments have relied on the "cheating game" (e.g. Mazar et al. 2008, Fischbacher & Föllmi-Heusi, 2013) which makes individual detection unlikely but possible. However, Kajackaite and Gneezy (2017) argue that eliminating all concerns for detection can in fact drive up cheating rates: indeed, when detection is made impossible through a modified version of the "mind game" (Jiang 2013, Shalvi & De Dreu 2014, Potters & Stoop 2016), Kajackaite and Gneezy find that higher stakes (ranging between $1 and $50 in a laboratory setting) do lead to an increase in cheating rates. Conversely, it is important to note that, unless detection by the experimenter is made completely impossible, social identity concerns are likely going to affect the intrinsically-perceived cost of lying (Gneezy, Kajackaite & Sobel 2017): participants who want to maintain a positive self-image are likely also concerned about their appearance towards a potential observer, including the experimenter, and as such increased cheating at a higher stake size might signal a negative image towards themselves and the potential observer.

In sum, social identity is an important element of self-image preservation: people's beliefs about being a moral or immoral person are shaped by whether others view them as moral or not, and these notions can become internalised. As such, people's self-perception of morality is, at least in part, socially constructed but also deeply internalised about what is right and wrong. This implies that it is unlikely that participants will cheat more at higher stake sizes relative to lower stake sizes because by doing so they would appear to be "a liar" both to themselves and others; meanwhile, cheating "just a little bit" (at smaller and higher stakes) may not have that same stigma and participants are thus likely to do so.

Taken together, these published findings suggest that people engage in at least some levels of cheating regardless of stake size, but that cheating rates do not increase with stake size.

H1: Participants in all conditions will cheat at least to some extent.

H2: Participants will cheat only a little more as the stakes increase.

1.2. Moral consistency and moral balancing

After engaging in a moral or immoral behaviour, two potential follow-on behaviours can occur. On the one hand, moral behaviour may be followed by more moral behaviour while immoral behaviour is followed by more immoral behaviour – a "consistency" account of morality (Zhong et al. 2009). Research on consistency has found that people behave consistently when they think their prior action was justified (Jordan & Monin 2008), want to reassure (and update their own beliefs about) themselves based on their prior actions (Ariely & Norton 2008), or want to avoid looking like a hypocrite (Cialdini et al. 1995). For example, the well-known Foot-in-the-Door paradigm shows that larger requests for one's time or commitment are more likely to be accepted when it was preceded by a smaller request (Freedman & Fraser 1966). This is in part driven by the fact that people derive some understanding of their own preferences based on what they observe themselves doing (Bem 1972, Ariely & Norton 2008).

Moral consistency usually occurs when the behaviour can be linked to an individual's identity concern. Specifically, Conway & Peetz (2012) demonstrate that, when people reflect what their actions imply for their self-image of a moral decision-maker, they are more likely to act morally, as compared to thinking about the specific actions in isolation. Furthermore, they also find that temporal distance from a particular action can affect the need to compensate behaviour or be consistent with it: behaviour that is morally good and further in the past is seen to reflect on one's identity and leads to consistency, but when a good action was taken only recently, moral compensation seems acceptable (Conway & Peetz 2012). Conversely, moral consistency can also be primed by products which are morally questionable and when the user is aware of the inauthenticity of the product: Gino, Norton, and Ariely (2010) find that participants who knew that they were wearing counterfeit sunglasses behaved less morally than those who knew they wore authentic sunglasses.

In sum, moral consistency is a possible consequence after taking an earlier moral action. Thus, in our setting, if consistency is expected, then those who cheat the most are also less likely to choose to donate and make smaller donations.

H3: Participants who cheat more are (i) less likely to give to charity and (ii) give a lower percentage of their earnings to charity.

On the other hand, moral behaviour at an earlier point in time could give license to less moral behaviour later, especially when the moral behaviour just recently preceded another moral action (Conway & Peetz 2012). Research on moral balancing (Miller & Effron 2010, Zhong et al. 2009) suggests that moral decisions are not viewed in isolation but usually within the context of previous and future decisions that help establish and maintain one's moral self-image. If the goal is to maintain a moral self-image overall, one need not act perfectly moral all the time: a moral decision today might invite an immoral decision tomorrow (moral licensing) and conversely an immoral decision might be followed by a moral one to make up for bad behaviour (moral cleansing).[7]

For example, Monin and Miller (2001) show that participants that have the chance to disagree with blatantly sexist statements are more likely to later hire a man for a stereotypically male job in a hypothetical hiring decision. Likewise, Effron et al. (2009) demonstrate that endorsing Obama can lead to racial discrimination in a later decision. Even hypothetical, counterfactual or potential thinking in the future can license other moral decisions (Effron et al. 2012, 2013, Gneezy et al. 2012). Conversely, knowing that an opportunity to "cleanse" immoral behaviour will exist in the future leads to more unethical behaviour early on (referred to as "conscience accounting," see Gneezy et al. 2014). Cojoc and Stojan (2014) also find that people behave more immorally if they are aware of an opportunity to behave morally in the near future. Further, they find that individuals who are aware of an opportunity to donate in a subsequent part of the experiment tend to give less than those who are not informed of that opportunity, independent of whether they behaved honestly or dishonestly in a cheating task.[8]

While the existence of moral licensing and cleansing across domains is well-established (for an introduction and overview, see Miller & Effron 2010), we know relatively little about the cognitive processes and psychological costs that moral balancing requires. Here we offer one account for when and how participants engage in moral balancing when stakes matter.[9] Resisting the temptation to cheat is cognitively taxing, as it requires self-control not to give in to cheating to one's benefit (e.g. Gino et al. 2011, Greene & Paxton, 2009). We argue that the temptation to cheat increases for at least some people

---

[7] An open question which is beyond the scope of this paper is to theorise about the dynamics and moderators that explain when moral consistency and moral balancing occur; for reviews and perspectives, see Dolan & Galizzi (2015) and Truelove et al. (2014).

[8] These findings may raise concerns in our study that the decision whether to cheat or not could be affected by knowledge of another moral task in the future. In our experimental design, however, participants did not know that they would later be asked to make a donation decision, so they could not have anticipated or adjusted their behaviour accordingly.

[9] In an earlier conception of this research, we hypothesised that our experimental setup would trigger moral cleansing (see link and abstract in the pre-registration in the online appendix): if participants cheat more in the coin-flipping task as the stakes increase, they might be more likely to donate a larger percentage to charity later, to "morally cleanse" for cheating at a higher rate. However, only later did we realise that the repeated mind-game was unlikely to lead to increased cheating in the coin-flipping task. Based on the fact that participants did *not* cheat in the first task, we then looked at the flip-side of moral balancing (comprising of moral cleansing and moral licensing) and predicted that this self-controlled behaviour at higher stakes would lead to more moral licensing when choosing how much to donate.

when the stakes for cheating rise, making it harder for them to resist the temptation. As a consequence, we predict that people who feel particularly virtuous and moral about their past behaviour—especially when the stakes are high—are more likely to feel justified later to behave more immorally. Indeed, the literature on moral licensing argues that people have an incentive and a tendency to view their benefits and costs in an asymmetric fashion (Miller & Effron, 2010). They are more likely to play up the good deeds they have done, even if just in their own minds, but downplay any negative behaviour. Due to this self-serving asymmetry, it is very likely that the higher the cheating benefits the participant resists, the higher their perception of their morality, leading to more licensing behaviour.

Assuming that participants exhibit some form of licensing behaviour after doing a good deed, past research suggests that there exist at least two possible explanations for how people might engage in moral licensing (Miller & Effron 2010). The "moral credits" model proposes that people keep an internal "moral balance sheet:" good deeds add to the moral balance while bad deeds are like debits on the balance sheet. This model suggests that people use their moral balance to "purchase" themselves the right to a moral transgression in the future (Hollander 1958, Merritt et al 2012). That is, even a blatant and unambiguous moral transgression is acceptable to a moral-credits decision-maker assuming that he or she has accumulated the moral credits previously to make up for it.

In contrast, the "moral credentials" model (Miller & Effron 2010, Effron et al. 2009) proposes that participants who behave morally in one decision are more likely to construe later moral transgressions as ambiguous and not immoral, having previously established that they are a moral decision-maker. Miller & Effron succinctly describe what a decision-maker might ask themselves: " 'Can I say or do this without signaling something morally discrediting about myself?' " (Miller & Effron, 2010: 119). Therefore a moral-credentials decision-maker would not view an ambiguous moral transgression as immoral because they have established credentials that show otherwise.

Both the "moral credits" and "moral credentials" models predict licensing after behaving morally but they differ in the pathway to getting there. In our experiment, participants make two subsequent decisions regarding giving to a charitable cause, each operating through a different pathway. Specifically, we ask participants if they would like to donate at all to charity and if so, how much of their earnings they would like to donate.

Participants choosing whether and how much to donate might display the two distinct pathways of moral licensing. The first—consistent with "moral credits"—implies that participants who have resisted cheating are *less likely* to donate anything to charity as stake size increases. Choosing not to give to charity is an unambiguous signal the decision-maker is sending after having previously behaved morally in the high stakes condition. A smaller fraction of donors as stake size increases is evidence for the "moral credits" account.

The second is consistent with the "moral credential" explanation: As stake size increases, participants give a *smaller proportion* of their earnings to charity. Donating a low percentage of a high earning is an ambiguous signal of licensing: when the stakes are high, participants who give a smaller percentage still donate a large absolute amount of their earnings to charity: as such they can construe their actions not to be immoral, thus behaving less morally without taking a hit to their moral self-image. As stake size increases, a lower percentage of earnings donated is evidence for the "moral credentials" account.

In sum, the two pathways to moral licensing make the following distinct but not mutually exclusive predictions about subsequent donation behaviour:

> *H4a ("Moral Credits" Model): Participants in the high stakes condition will be less likely to donate to a charity.*
>
> *H4b ("Moral Credentials" Model): Participants in the high stakes condition will give a smaller percentage of their earnings to charity.*

1.3. Self-reported morality

Moral balancing is simultaneously an internal, cognitive process as well as an outward projection: participants continuously evaluate their actions against a backdrop of previous behaviours and situations and justify the morality of their actions to themselves and others (Zhong et al. 2009, Jordan et al. 2011, Shalvi et al. 2011, 2012). Thus, we are not just interested in measuring behaviour such as cheating or donations, as perceived by outside spectators, but also in participants' self-reports of their own morality – that is, how participants perceive their (im)moral behaviour, and how those self-reports might predict their future moral behaviour.

Participant morality relating to a situation may be assessed by implicit or explicit measures. Implicit measures include solving word fragments with terms which may or may not relate to morality. For example, Gino et al. (2011) used word fragment puzzles which could be solved with morality-related terms such as 'moral', 'virtue' or unrelated terms such as 'mural' or 'tissue.' Implicit elicitation of morality is useful in circumstances, where, for example, concerns of social desirability prevail, or moral awareness may not be activated (Bazerman & Banaji 2004, Rudman 2004).

In contrast, explicit self-reports of morality involve asking a participant how they feel or anticipate feeling in regard to a moral dilemma (e.g., having the opportunity to cheat on a coin-flipping task). Self-reports are helpful in understanding how a decision-maker views themselves, and portrays themselves to others, after acting (un)ethically (Rudman 2004). Effron et al. (2015) ask participants, for example, how guilty, unethical, dishonest they feel after such a task. Similarly, Zhong and Liljenquist (2006) elicit self-reports of moral emotions after engaging in a physical cleansing task. Gino and Desai (2012) use 'moral purity' to assess how 'innocent' and 'morally pure' participants felt.

We predicted that the higher stakes of cheating in some conditions would lead to more temptation to cheat, which requires effort to resist but which also provides participants with an opportunity to claim and leverage virtuosity. We therefore expect that self-reports of morality would provide a clue to understanding the presence of moral licensing. Specifically, resisting temptation to cheat when the stakes are high might lead some participants to feel highly moral, which they leverage to justify their subsequent immoral behaviour. We consider two outcomes to measure (*i*) an effect on self-reported morality on average as stakes increase and (*ii*) a tendency to leverage perceptions of high morality to excuse less moral behaviour subsequently.

> *H5a: Self-reported morality increases with higher temptation to cheat.*
>
> *H5b: Those who perceive themselves as highly moral in the high stakes condition give a smaller percentage of their earnings to charity.*

1.4. Reflecting on morality

We have argued that resisting the temptation to behave dishonestly could lead to unexpected downstream consequences in a subsequent moral decision and that self-perceptions of morality can be used to justify immoral behaviour. But what happens after the 'heat' of the cheating moment has passed?

Empirical research on the management of a positive self-view after unethical behaviour (Chugh & Kern 2016), and in particular reflections on unethical behaviour to manage a positive self-view, has received little scholarly attention. Only recently, Kouchaki and Gino (2016) proposed a pathway through which people might deal with their unethical past. The authors give participants the opportunity to cheat a little. Several days later, they invite the same participants back to the study and ask them to remember the details of the cheating (versus non-cheating) task. They find that participants obfuscate their past unethical behaviour by forgetting the details of the cheating task they had to complete. They engage in "unethical amnesia" – the subconscious and purposeful forgetting of details of their past unethical acts (Kouchaki & Gino 2016).

Here we investigate reflections on feelings of morality—not the factual details of the task—one day after the completion of a cheating task. Kouchaki and Gino's research suggests that those who cheat a little engage in obfuscation of their actions and do not report feeling morally different than those who behaved completely honestly.

However, we propose that obfuscation of the memory of feeling immoral does not apply to "obvious cheaters" who are in clear violation of the rules: when behaviour is unambiguously immoral (such that there is no "moral wiggle room" (e.g., Dana et al. 2007) to pretend otherwise), feelings of morality will—perhaps surprisingly—be lower than those who cheat only a little and, importantly, become worse over time. Specifically, we propose:

> *H6a: Participants who cheat a little feel morally (i) no different than those who behave completely honestly and (ii) no different upon reflection a day later.*

> *H6b: Participants who cheat maximally, such that their behaviour is unambiguously immoral, will feel (i) morally worse than those who do not cheat or cheat only a little and (ii) feel even worse when reflecting on their morality a day later.*

Finally, as the stakes increase, we would expect this effect to increase, such that the overall effect of maximal cheaters who feel guiltier as time passes arises primarily from the high stakes conditions, where maximal cheating yields the greatest payoff.

> *H6c: The effect of maximal cheaters feeling worse upon reflection is strongest in the highest stakes condition.*

2. Experimental design

2.1. Experimental design overview

Our experiment was conducted on Amazon Mechanical Turk (MTurk)[10] over two days in January 2017

[10] Amazon Mechnical Turk is an online labour market which has been used extensively for economic research; however, it does not come without limitations and potential selection effects. We refer interested readers to Clifford, Jewell & Waggoner (2015), Horton, Rand & Zeckhauser (2011) and Landers & Behrend (2015) for a discussion of the limitations and opportunities of the MTurk platform for research purposes.

(see **Figure 1**). On the first day, we recruited a total of *N*=2,015 participants across four conditions. Participants received $1.00 for participating in the study and were told that they had the possibility to earn a bonus payment based on the decisions they made during the study. Since the maximum possible bonus varied by condition, the exact amount of potential bonus earnings was not disclosed at the time of recruitment.

In all conditions, participants first read the instructions for a coin-flipping task. All participants engaged in 10 rounds of the coin-flipping task in which they had the opportunity to cheat to earn an additional bonus payoff. We randomly allocated participants to one of four conditions, which varied the maximum potential payoff that they could earn from the coin-flipping task ($0.10, $0.50, $5.00, or $50.00). All payoff-relevant decisions were incentive-compatible.

After the coin-flipping task, participants self-reported their feelings of morality and indicated what percentage (if any) of their earnings they want to give to a charity of their choice. The participants finished the study by responding to a few vignettes regarding guilt proneness and filling out demographic information.

On the second day, we invited all participants back to complete a short survey. A total of *N*=1,413 participants (70%) returned for the second survey.[11] In the second survey, participants reported whether they won in the lottery draw and again self-reported their feelings of morality from the coin-flipping task completed a day earlier.

2.2. Measuring dishonesty

Our experiment involves participants playing a variant on the traditional cheating game introduced by Jiang (2013). Tasks that involve flipping a coin or rolling a die in private are unobtrusive measures of honesty (e.g. Bucciol and Piovesan, 2011). Traditional coin-flipping tasks reward the participant based on a pre-determined outcome, as specified by the experimenter (e.g., the participant gets paid if the coin

[11] We tested for differential uptake of the second survey. We did not find any correlation between uptake and cheating behaviour in the coin-flipping task, self-reported morality, donation likelihood or donation level using logit models (using logit to predict uptake of the second survey by each of the independent variables: all *p*s > 0.05). There were, however, some differences in uptake across conditions (Kruskal-Wallis = 8.23, $p = 0.042$) though trend effects from higher stakes were only marginally significant (Jonckheere-Terpstra = 778240, $p = 0.072$). In conducting pairwise comparisons between the conditions, we used bonferroni-corrected *p*-values. The only significant difference was between the $5 and $50 maximal payoff conditions (Wilcoxon = 116780, $p = 0.03$, bonferroni-corrected) such that more participants in the $50 condition completed the second survey. All other post-hoc comparisons were not significant (*p*s > 0.1).

comes up heads). The participant flips the coin in private and then reports the outcome. While this task has been shown to lead to moderate rates of cheating, it has been argued that it does not conceal a participant's dishonesty completely because the participant might think that the experimenter would be able to detect their dishonesty if they are able to check the actual outcome of the coin flip.

The 'mind game' variant of the traditional cheating game enables a more robust concealment of dishonesty from the participant's perspective. Specifically, in our design, participants are asked to (*i*) think of an outcome from a coin toss and remember it, (*ii*) toss a coin in private and report the outcome, and (*iii*) report whether the actual outcome matched the outcome they had thought of. In this setup, the experimenter is unable to verify, on an individual level, whether the reported outcome matched what the participant thought of previously. Participants play 10 rounds of the same coin-flipping mind game. Because the participant knows that the experimenter cannot access his or her thoughts, the participant can be assured that he or she will not be detected. Indeed, Kajackaite & Gneezy (2017) have demonstrated that cheating rates in a (single-round) mind game increase at higher stakes compared to the equivalent traditional cheating game. While the experimenter cannot identify on an individual level which participant cheated, it can be determined whether cheating occurred at a group level based on the theoretically expected probabilities of matches occurring in the group. The average cheating rates can thus be compared across conditions to detect whether more or less cheating occurred as incentives changed.

Participants' potential earnings depended on the number of reported coin flips matching their imagined outcome. We varied the incentive to cheat across four conditions: in condition 1, participants earned US $0.01 per matching coin flip outcome; in condition 2, US $0.05 per match; in condition 3, US $0.50 per match; and in condition 4, US $5.00 per match. Across all 10 rounds of the game, participants could therefore earn a maximum of US $0.10, US $0.50, US $5.00 and US $50.00 in conditions 1 through 4, respectively. The condition with a maximum potential earning of US $50.00, which we will refer to as "high stakes" condition, represents significant stakes for crowd-sourcing platforms like MTurk where reservation wages of US $1.4 dollars per hour have been measured (Horton and Chilton 2010).

A competitive lottery mechanism was used to award earnings. The lottery approach was used to manage the costs of running the experiment and has been found not to distort true preferences (Starmer and Sugden 1991). In particular, following the procedure by Cohn et al (2014), participants were told that their number of coin-flip matches would be compared to another randomly selected participant; if the participant reported an equal or higher number of coin flips to the comparator participant, they were entered in the lottery draw. In the lottery, one in five subjects are randomly selected as winners and received their earnings as a cash bonus payment on the second day of the experiment.

2.3. Measuring self-reported morality

We used self-reports of morality to measure the costs of cheating, or resisting the temptation to cheat, on both day 1 and 2 of the experiment. After completing the mind game coin-flipping task, we asked subjects to complete self-reports on how moral, virtuous, dishonest and unethical they felt on a 5-point Likert scale (where 1 = not at all, slightly, somewhat, very much, 5 = extremely). The adjectives were presented in a randomised order, and a combined 'morality' index for analysis was created.

The same four morality items were assessed again in the follow-up study. A day later after the coin-flipping experiment, we invited participants to answer the same questions about their moral feelings in relation to the coin-flipping task they completed a day earlier.

2.4. Donation behaviour

After completing the coin-flipping task and reporting their morality on day 1 of the experiment, participants were asked if they would like to donate any of the potential earnings from the coin-flipping task to a charity of their choice. Participants were told they could choose from a list of six popular US charities.[12] If participants chose to answer the first question—"would you like to donate some or all of the bonus to a charity?"—with "yes," they were presented with two follow-up questions: (*ii*) "Which charity would you like to donate some or all of any bonus awarded?" and (*iii*) "What percentage (%) of any bonus awarded would you like to donate to your chosen charity?" To enable comparability across conditions, we asked what proportion of their potential earnings they would like to donate (0-100%). The slider scale was anchored at 50% and increments of 25% were available.[13]

2.5. Control and exploratory variables

At the end of the survey, participants filled out a short survey to control for individual variation and to allow for exploratory analysis. The items included the 'negative behavioural evaluation' subscale of the Guilt and Shame Proneness (GASP) scales, materialism, competitiveness (Cohen et al. 2011). In addition,

[12] The charities to choose from were the Red Cross, St Jude Children's Research Hospital, Salvation Army, United Nations Children's Fund (UNICEF), American Society for the Prevention of Cruelty to Animals (ASPCA), Habitat for Humanity. We selected these charities based on their popularity and brand recognition.
[13] The survey software used, Qualtrics, requires that the slider scale have an anchor.

we collected information on demographics, past donation behaviour and frequency, and MTurk experience both in terms of years and past participation in coin-flipping tasks.

2.6. Pre-test experiments and power analysis

Before conducting the main experiment we ran a pre-test experiment (*N* = 180). The pre-test was conducted for three reasons. First, we aimed to pre-test and correlate all stated measures in a non-overlapping sample of MTurk participants with actual cheating behaviour. Specifically, we explored implicit and explicit approaches to measuring morality: after the "mind game," participants were randomly assigned to either solving word fragments with terms relating to cheating costs (e.g. b_d -> bad, _ _ eater -> cheater)[14] or self-reported measures of feeling virtuous, moral, dishonest and unethical (as described in Section 2.3), respectively. We adopted aspects of the self-report scales from Effron et al. (2015), which measured guilt and virtue after a cheating task.

We were interested whether actual cheating behaviour over 10 rounds of the same "mind game" as in the main experiment (as described in Section 2.2) affected morality, as measured implicitly by the word fragment puzzle or explicitly through self-reports of moral emotions (or both). Unlike the main experiment, however, there was no randomisation of stake size; for all participants, the total stake size across the ten rounds was fixed at $0.50 (i.e., the equivalent of condition 2). We found significant positive correlations between the matches reported in a cheating task and self-reports of feeling dishonest and unethical (rho = 0.35, $p = 0.003$). However, we did not find a significant relationship between matches reported and the other outcomes, including morality-related words solved for in the word fragment task (all *p*s > 0.05). We thus decided to retain the self-reported measures and exclude the word fragment task in the main experiment.

Second, we aimed to determine the minimum sample size needed to detect small changes in our main outcome variables in the main study. For conservative measure, we focused on a subset of participants in the second pre-test experiment who completed the moral self-reports and reported not knowing the purpose of the coin-flipping task. We simulated 1,000 random samples for differently-sized groups, and determined the proportion of samples for each group that found a significant positive correlation between the number of coin flips reported as matched and negative moral affect using a one-sided Spearman test at the 5% significance level. Using this procedure, we determined that a sample size of at least 460 participants per condition was required to achieve 80% power in the main experiment. We decided to

[14] Gino et al. (2011) use this method to measure access to ethics-related words.

round up to 500 participants per condition to further increase power.

Finally, we were interested in testing whether including the self-reported morality scales before the donation decision would affect this subsequent moral decision. We found that neither measure of morality (self-reports and solving word fragments) significantly affected the subsequent likelihood of engaging in a charitable donation nor the proportion of earnings to be donated (both *p*s > 0.05). We thus decided to keep the morality measures between the cheating and donation tasks for the main experiment.

3. Results

3.1. Dishonesty: little cheating for big money

We begin by looking at the effects of stake size on cheating behaviour in the coin-flipping task. First, we look at the average number of times participants in our experiment reported a 'match.' Given sufficient number of observations, the expected theoretical value from ten coin-flips per participant is 5 matches if we assume all participants act entirely honestly. However, the actual number of reported matches in our dataset is 6.28, above the expected value. Using a two-sided one-sample *t*-test, we reject the null hypothesis that the expected value equals the actual reported average number of matches per participant, $t(2{,}014) = 31.94$, $p < 0.001$. Furthermore, we conduct a nonparametric comparison between the theoretical distribution expected under full honesty (i.e., the theoretical distribution of a fair coin) and the observed empirical distribution of matches reported: we detect a significant difference between the distributions ($Z = 2{,}817{,}200$, $p < 0.001$), suggesting the presence of cheating. Thus, in line with our prediction in H1, we find significant levels of low-level cheating in our sample.

Second, we compare cheating behaviour across the four conditions with varying incentives to cheat per coin-flip. We find that increasing stake size leads to significantly more cheating but the differences between smaller and larger stakes are economically small (**Figure 2A**); the distributions of matches reported across 10 rounds are qualitatively similar (**Figures 2B-E**). Formally, we find significant differences between the distributions of matches across conditions using a Kruskal-Wallis test ($H(3) = 17.884$, $p < 0.001$). We also conduct a Jonckheere–Terpstra test to compare the ordering of the number of matches reported across the four conditions and find a significant effect of increasing stake size on the level of cheating, ($T_{JT} = 810{,}720$, $p < 0.001$). Bonferroni-corrected post-hoc pairwise comparisons with Wilcoxon tests show that there were significant differences between the \$0.10 and \$5.00 condition ($W = 107{,}900$, $p < 0.001$), \$0.10 and \$50.00 condition ($W = 113{,}640$, $p = 0.005$), and the \$0.50 and \$5.00

condition (W = 115,140, p = 0.007). However, there were no significant differences between the two low-stakes condition ($0.10 and $0.50, $p > 0.10$) and the two high-stakes condition ($5.00 and $50.00, $p > 0.10$).

These results suggest that a 500-fold increase in the size of stake to cheat increases cheating only very little. To quantify the extent of cheating by stake size, we predict average matches with a continuous variable measuring the increases in stake size. Fitting a linear regression using a continuous variable with the four stake sizes in cents (1, 5, 50, and 500) reveals only a weak, marginally significant relationship with cheating behaviour (coeff = 0.0003, p = 0.101; see Online Appendix **Table A1**). A log-transformation of the stake size, instead, is a better predictor in the same model (coeff = 0.055, p = 0.001), suggesting that each 10-fold increase in stake size leads to reporting only a 0.05 higher number of average matches reported (**Table A1**).

Taken together, these results support our second hypothesis (H2) that, while cheating does significantly differ across conditions, the effect sizes are economically small, such that participants are generally insensitive to stake size.

3.2. Moral consistency and balancing

3.2.1. Smaller percentage donated at higher stakes

We next focus on understanding the downstream consequences of cheating, or resisting the temptation to cheat, on future moral behaviour. We first hypothesised that those who cheat more are also generally less likely to give to charity and give less to it. Conversely, however, for some participants resisting the temptation to cheat will be psychologically taxing the larger the stakes are, and consequently, we predicted that higher stakes can trigger moral licensing in a subsequent moral decision.

To examine the potential consistency and balancing response, we look at the effect of stake size on the probability to make a donation and the donation amount. Using these two outcomes, we are able to distinguish between the "moral credits"—measured by the likelihood to donate—and "moral credentials"—measured by the percentage of earnings donated—explanations for licensing.

First, we observe that the probability to make a donation does not differ by condition (using logit regression predicting making a donation of any amount by condition dummies: all ps > 0.1, **Table A2**)

(**Figure 3A**). Furthermore, we also do not find that those who are more likely to have cheated (i.e., higher number of matches reported) are any more (or less) likely in the high stakes conditions to donate (using donation likelihood predicted by interaction of condition and total number of matches: all *p*s > 0.1), though total number of matches independently predicts lower likelihood to donate across conditions (without interaction: coeff = -0.150, $p < 0.001$; with interaction terms: coeff = -0.179, $p = 0.003$). This suggests that (*i*) those who cheat more are less likely to give to charity, lending support that at least some of our participants are engaging in moral consistency (H3) and (*ii*) there is no evidence of blatant moral licensing at higher stakes in the form of lower likelihood to donate, suggesting that these results do not provide support for the "moral credits" model (H4a).

Conversely, the "moral credentials" model suggests that stake size might affect donation levels, a more ambiguous behaviour to license. The donation level is the percentage of earnings from the coin-flipping task that a participant chooses to donate, provided they selected to make a donation in the previous question. We find that donation levels do vary significantly with condition (**Figure 3B**): relative to the $0.10 condition, participants in the $5.00 condition (coeff = -13.403, $p < 0.001$, **Table 1**) and $50.00 condition (coeff = -19.850, $p < 0.001$) donate significantly less to charity but not in the $0.50 condition ($p > 0.1$). All other pairwise differences between conditions are significant (*p*s < 0.05).

These results remain qualitatively robust to including the number of matches as a covariate: participants in both the $5.00 and $50.00 conditions donate significantly lower percentages of their earnings to charity (*p*s < 0.001). Conversely, the coefficient on the number of matches reported is significantly negative, giving support to the notion that some participants were exhibiting moral consistency (H3; coeff = -3.300, $p < 0.001$). Finally, we consider interactions with the number of matches reported, and find that the treatment effect is not dependent on the likelihood to cheat ($p > 0.1$ for all interaction terms). That is, those who are more likely to have cheated do not donate any more or less in the high stakes conditions. Taken together, these results provide support for the predictions of the "moral credentials" model (H4b).

3.2.2. Leveraging high morality to justify lower donations

To understand how participants cognitively process giving less to charity, we explore participants' self-reported morality. Morality was assessed immediately after, and in reference to, the coin-flipping task. We first look at average levels of morality across conditions, followed by individual-level correlations between moral self-reports and donation behaviour.

Across all conditions, we do not find significant differences in the average levels of morality reported (**Table 2**; using linear regression predicting self-reported morality by condition dummies; followed by pairwise comparisons of coefficients: all *p*s > 0.1). There are also no interactions with the behaviour in the coin-flipping task (using linear regression with condition dummies interacted with number of total matches reported: all *p*s > 0.1), though the total number of matches reported in the coin-flipping task does negatively predict morality (without interaction terms: coeff = -0.076, $p < 0.001$; with interactions: coeff of single effect = -0.052, $p = 0.008$), suggesting that participants' self-perceptions of their morality, on average, correlate with actual cheating behaviour in the coin-flipping task. These results thus do not provide evidence that participants on average inflate their morality when resisting the temptation of greater stakes (H5a).

However, the aggregate view of morality on the condition-level might obfuscate individual-level cognitive processes. We argued that resisting the temptation to cheat in the high stakes condition can be used by participants to justify their licensing behaviour (i.e., donating a smaller percentage of their earnings). That is, while participants may, on average, not believe they are more moral in the high stakes condition, some participants could construe their resistance to not give in to cheating when the stakes are high as a justification for deserving more for themselves in a future moral decision. If so, we would expect a negative correlation between participants' self-reported morality and the percentage they donate to charity in the high stakes condition, but not in other conditions.

This is exactly what we find: donors in the $50.00 condition who feel *more* moral donate *less* to charity (using linear regression predicting donation level in the $50.00 condition by morality: coeff = -6.425, $p = 0.009$, **Table 3**), while the morality of participants in lower-stakes condition does not predict donation levels ($p > 0.1$, except in the $0.10 condition where, in fact, donors give marginally *more* to charity the higher the self-reported morality, $p = 0.114$, suggesting that at the lowest stakes this relationship might point towards moral consistency rather than licensing). In addition, when all interaction terms between conditions and morality are included in the regression, the interaction between morality and the high stakes condition is significant (predicting donation level by condition and morality: coeff of interaction term of $50.00 condition = -11.839, $p = 0.007$; all other interaction terms, *p*s > 0.1). These results are further robust to including the number of matches reported.

In sum, while there is no evidence for inflation of morality in the high stakes condition generally, we found support for the hypothesis (H5b) that participants in the high stakes condition who claim they are particularly moral donate a smaller percentage of their earnings to charity.

3.3. Moral self-perceptions

3.3.1. Reflecting on morality makes maximal cheaters feel morally worse

Finally, we turn to investigating self-perceptions of one's morality. We begin by comparing self-reported morality immediately after the cheating decision and one day afterwards. Overall we find that participants report lower levels of morality one day after (M = 4.123, *s.d.* = 0.766) than immediately after completing the cheating task (M = 4.172, s.d. = 0.727), $t(1412) = 3.239$, $p = 0.001$.

However, specifically, we predicted that those who cheat a little would not feel different than those who behaved honestly and neither of those groups would feel different a day later. Conversely, those who cheated a lot would feel less moral in general and more so one day later.

To test these predictions, we group participants by the number of matches they reported during the coin-flipping task. We take as baseline the group of participants who reported between 0 and 5 matches: in this range participants "almost certainly" behaved honestly – i.e., they are statistically likely to have reported the true outcome of the coin flip.[15] We created three additional groups based on the number of matches reported, in increasing likelihood of cheating across the ten round of the coin-flipping task: group 2 contains all participants who reported 6 or 7 matches ("likely honest"), group 3 is made up of participants who reported 8 or 9 matches ("likely dishonest"), and group 4 consists of participants who reported 10 matches. We have referred to the latter as "maximal cheaters" above, given that they have "almost certainly" behaved dishonestly in at least some round of the coin-flipping task.

Consistent with our predictions (H6a), we find no large differences in morality between groups 1, 2, and 3 immediately after the cheating task (using linear regression predicting morality by group dummies: group 1 vs. 2: $p > 0.1$; group 1 vs. 3 and group 2 vs. 3: $p = 0.071$ and $p = 0.075$, respectively, suggesting a marginal decrease in morality as the likelihood of that the participants cheated increases; **Table A3**). Furthermore, participants in these three groups did not report a change in their morality one day later (using *t*-tests for each group, testing morality ratings immediately after the task with morality ratings one day later, all *p*s > 0.1).

[15] In theory, participants could of course have lied and misreported the outcome of any of the single round of the game, which we cannot detect given that guessing the correct outcome of the coin-flips 50% of the time is the most likely outcome. For the purposes of the discussion here, we will refer to this group as "almost certainly honest."

In contrast, maximal cheaters reported significantly lower morality ratings than any other group (using linear regression predicting morality by group dummies: group 1 vs. 4: coeff = -0.693, $p < 0.001$; group 2 vs. 4: coeff = -0.689, $p < 0.001$; group 3 vs. 4: coeff = -0.597, $p < 0.001$, **Table A3**).[16] Moreover, those participants also reported feeling less moral than a day earlier when reflecting back on the coin-flipping task, $t(117) = 3.897$, $p < 0.001$. In fact, in support of our predictions (H6b), the group of maximal cheaters was the only group that reported reduced feelings of morality on the second day (**Figure 3A**; using linear regression predicting difference in self-reported morality between day 1 and 2 by group dummies, see **Table A4**).

What role did stake size play for self-perceptions? While we expected self-reported morality to be especially low among maximal cheaters in the high-stakes condition, we do not find strong evidence for this prediction: when the difference in self-reported morality immediately after the coin-flipping task and one day later is regressed on group dummies independently for each condition, we find that this difference measure for maximal cheaters is significantly lower, as expected, in the higher-stakes conditions ($5.00 stakes: coeff = -0.370, $p = 0.026$; $50.00 stakes: coeff = -0.350, $p = 0.031$) than the $0.10 baseline, whereas there is no significant difference between the lower stakes conditions ($0.10 stakes: coeff = -0.222, $p > 0.5$; $0.50 stakes: coeff = -0.100, $p > 0.1$). While this analysis might at first suggest that stake size do play a role, we do not find significant differences when all interaction terms are included in the full regression (all *p*s > 0.1, **Table A4**); nor when we compare the full interaction of group and stake size on self-reported morality on the first day (all *p*s > 0.1, **Table A3**).

We conclude that there is only limited support for H6c that stake size has an effect on self-perceptions; instead, maximal cheaters, regardless of stake size, feel worse than everyone else immediately after engaging in cheating behaviour, and worse as time passes.

3.3.2. Exploratory analysis: misprediction among maximal cheaters

Finally, we turn to an additional analysis using variables which we included for exploratory purposes. While we did not *a priori* hypothesise the relationships reported below based on past literature and theory, we believe they show promising insights for future research.

Based on the fact that maximal cheaters were more likely than any other group of participants to feel

[16] Results are qualitatively similar when we control for general guilt proneness in these regressions.

worse one day later (Section 3.3.1), we became interested in understanding whether participants were able to foresee this negative change in self-perceived morality. To find out, we looked at the 'negative behavioural evaluation' subscale of the Guilt and Shame Proneness (GASP) scale. This scale captures the tendency that people would feel uncomfortable and guilty after committing an unethical act, with higher scores indicating feeling more remorse after behaving dishonestly. We refer to this scale as "guilt proneness."

Participants in general expressed that they would feel relatively uncomfortable if they committed an unethical act (M = 5.430, *s.d.* = 1.354). However, there exists considerable variation across participants based on their behaviour: participants who reported a higher number of matches in coin-flipping task expressed lower guilt proneness (using linear regression predicting the GASP guilt measure by total number of matches: coeff = -0.115, $p < 0.001$).

Perhaps surprisingly, participants who cheated maximally expressed the lowest guilt proneness compared to any group based on matches reported (linear regression predicting guilt proneness by group dummies, followed by pairwise comparisons: all *p*s < 0.05; **Table A5** and **Figure 4B**). However, as reported in the previous section, maximal cheaters were also the only group that did, in fact, show a negative change in feelings of morality one day after the task.

Taken together, these results suggest that maximal cheaters were not very good at predicting their own moral feelings: although they claimed that they would generally not feel guilty after behaving unethically, they were the only group of participants that showed a *decrease* in self-reported morality one day after committing an unethical act.

4. Discussion

We have shown that cheating only occurred at low levels, even when the stakes were extremely high for the online participant sample in this study. Yet, resisting the temptation to cheat at high stakes did have downstream effects on other moral behaviour, such as donating to charity. Specifically, we observed that participants gave a smaller fraction of their earnings to charity as the stake size increased. This suggests behaviour consistent with moral licensing: participants who refrained from cheating at higher stakes seem to have subsequently licensed themselves to donate less to charity, thereby "balancing" their moral behaviour over time. Indeed, we find that donors in the high-stakes condition who reported greater feelings of morality gave a smaller fraction of their earnings to charity. Finally, we observed a drop in self-reported feelings of morality one day after the task among maximal cheaters—but no other group of participants. This is an effect that the maximal cheaters did not appear to foresee, as they believed they were the least guilt prone after cheating.

The insensitivity to stake size we observed in our study is consistent with past work (Mazar et al. 2008, Abeler et al. 2016). In our setting, the maximum potential payoff in the lowest-stakes condition was $0.10 and in the highest-stakes condition $50.00 for a participant who was willing to cheat maximally across all rounds in the game. This marks a 500-fold increase in incentives. Given the profile of Amazon Mechanical Turk workers whose median reservation wages has been estimated as low as $1.4 per hour (Horton & Chilton 2010), the highest payoff amount in the high-stakes condition was likely an attractive incentive for many participants. Yet, a 10-fold increase in stake size resulted only in a 0.05 increase in reported coin matches. This finding is also echoed by a recent meta-analysis: Abeler et al. (2016) re-analysed the full datasets from 46 studies across 43 countries, finding almost no effect of higher stakes across a large range of stake sizes.

However, past work has also found circumstances when cheating is sensitive to incentives. Kajackaite and Gneezy (2017) demonstrate that, when individual-level detection by the experimenter is made completely impossible, participants do cheat at higher rates when stake size increases. While we employ a similar method to Kajackaite and Gneezy (2017) and Jiang (2013), the sense of not being able to be detected is likely somewhat diminished. Participants in our experiment played several rounds of the mind game, increasing the ability of an experimenter to infer cheating behaviour at an individual level. Still, while the experimenter cannot be sure if someone cheated, participants might be less willing to engage in blatant cheating. Consequently, we observe less cheating in the multi-round mind game than past work where detection was impossible in a single-shot game.

When the stakes in the cheating task in our experiment increased, participants subsequently gave a smaller fraction of their earnings to charity. Moral licensing theory proposes that people can engage in dishonest and selfish behaviour without incurring a cost to their moral self (Miller & Effron 2010). Our results fit a “moral credentials” explanation (Effron et al. 2009): participants leveraged their past moral credentials (i.e., resisting the temptation to cheat) to justify keeping more of their high earnings from themselves, while at the same time not seeming immoral. Merritt et al. (2010) argue that moral credentials work because the licensed action is ambiguous – immoral behaviour can be “reframed” to still be construed as moral, both to the decision-maker and others. Although participants were not aware of other conditions, participants in the higher-stakes conditions likely had an opportunity to reframe their charitable contributions in light of the fact that the amount they gave to charity (relative to typical charitable donations on Amazon Mechanical Turk) could be construed as quite high. That is, although they gave a smaller fraction of their earnings to charity, they nonetheless gave a larger absolute amount to charity than those in the smaller-stakes conditions – arguably a generous gift. Whether or not this is interpreted as morally permissible might depend on the point of view: Hauser et al. (2016) show, for example, that most participants in a group believe that group members with larger earnings ought to contribute the same (or a higher) percentage of their income to a public good, while participants with high endowments themselves think that giving a larger absolute—but not relative—amount suffices.

Overall, participants viewed themselves quite differently depending on their cheating behaviour. While a little cheating did not impact self-view of morality, participants who cheated maximally, independent of stake size, felt significantly less moral after the task. First, participants who cheat a little but do not feel immoral might engage in a mild form of self-deception. Von Hippel & Trivers (2011) argue that people engage in self-deception, often to further their own goals without paying the cost of feeling immoral – a process that Batson & Thompson (2001) call "moral hypocrisy." Self-deception may be aided by moral disengagement and motivated forgetting (Kouchaki & Gino 2016, Shu, Gino & Bazerman 2011, Bandura 1999). Yet, maximal cheaters in our experiment did not attempt to deceive themselves (or others) about the morality of their actions. But, while maximal cheaters acknowledged their immoral behaviour in the moment, they seem to have misjudged the cost on their moral self-view in the future: they were the only group of participants who felt less moral a day later and, at the same time, they believed they would not feel guilty after engaging in immoral behaviour. Chance et al. (2011) demonstrate that participants consistently mispredict self-deceptive behaviour, even when it comes at a personal cost. Our results extend this line of reasoning to self-perceptions of ethical decisions more generally: even when participants believe they are fine with behaving unethically today, their view of their own actions suffers at a later date. However, we did not explicitly measure prediction of future moral self-perceptions, a task which we encourage for future research.

Our experiment is of course not without limitations. For example, the variation in donation levels we observe may in part be due to a "wealth effect," as the differences in any small or large amount earned from the cheating task could potentially affect subsequent behaviour. However, previous studies have shown that donation decisions similar to the one we employed are largely unaffected by endowment or stake size. Most relevant to our investigation, Raihani et al. (2013) study large stake size variation in the dictator game and find no wealth effects in the same Amazon Mechanical Turk population we use here. Similarly, findings from variation in endowments in an ultimatum game by Andersen et al. (2011) suggest that variation in windfall stake size would likely not explain more than 30% of the effect. While we cannot rule out wealth effects in our setting (or that our results may not be affected by extremely large stake sizes, effectively placing a boundary condition on the proposed behaviours), it is unlikely that they play a significant role in our study based on these prior findings.

Our design and findings have the following important practical implications for organisations. First, while offering a high-stakes opportunity to cheat does not translate into meaningfully higher cheating rates, organisations should carefully examine the surrounding, temporal decision context: if employees and managers feel especially morally virtuous after a high-stakes moral decision, they might engage in less moral behaviour subsequently. As such, strategies that simply deter dishonesty without eliminating temptation (e.g. by increasing monitoring or punishment) may be insufficient, as they might inadvertently give a moral license. This can consequently reduce other pro-social behaviour (e.g. costly contribution to team projects, or resources directed at corporate social responsibility). Second, understanding these "spill-over" effects has implications for the design of organisational decision-making: since managers and corporate boards often face high-stakes decisions, they may be especially likely to "balance" their moral decisions – thus, a prudent intervention would assign multiple high-stakes decisions to different decision-makers. Third, when unethical behaviour coincides with a self-serving payoff, managers may subconsciously engage in some form of self-deception or downplay the likelihood that they would regret their choice. Future research could investigate whether prompting decision-makers with an intervention that informs them of the "spill-over costs" (i.e., that they might feel worse about their decision tomorrow) or that reminds them of their "moral identity" (e.g., linking it to their consistently pro-social behaviour in the past via "identity nudges," see Kessler & Milkman 2016) prompts momentary reflection and more ethical decision-making.

References

Abeler, J., Nosenzo, D., & Raymond, C. (2016). *Preferences for truth-telling.*

Andersen, S., Ertaç, S., Gneezy, U., Hoffman, M., & List, J. A. (2011). Stakes matter in ultimatum games. *American Economic Review*, 101(7), 3427-39.

Ariely, D., & Norton, M. I. (2008). How actions create–not just reveal–preferences. *Trends in cognitive sciences*, *12*(1), 13-16.

Bandura, A. (1999). Moral disengagement in the perpetration of inhumanities. *Personality and social psychology review*, *3*(3), 193-209.

Batson, C. D., & Thompson, E. R. (2001). Why don't moral people act morally? Motivational considerations. *Current directions in psychological science*, *10*(2), 54-57.

Bazerman, M. H., & Banaji, M. R. (2004). The social psychology of ordinary ethical failures. *Social Justice Research*, *17*(2), 111-115.

Bazerman, M. H., & Gino, F. (2012). Behavioural ethics: Toward a deeper understanding of moral judgment and dishonesty. *Annual Review of Law and Social Science*, *8*, 85-104.

Becker, G. S. (1968). Crime and punishment: An economic approach. In *The economic dimensions of crime* (pp. 13-68). Palgrave Macmillan, London.

Bem, D. J. (1972). Self-perception Theory in *Advances in experimental social psychology*, 6, 1-62. Academic Press.

Bucciol, A., & Piovesan, M. (2011). Luck or cheating? A field experiment on honesty with children. *Journal of Economic Psychology*, *32*(1), 73-78.

Cappelen, A. W., Sørensen, E. Ø., & Tungodden, B. (2013). When do we lie?. *Journal of economic behaviour and organization*, 93, 258-265.

Chance, Z., Norton, M. I., Gino, F., & Ariely, D. (2011). Temporal view of the costs and benefits of self-deception. *Proceedings of the National Academy of Sciences*, *108*(Supplement 3), 15655-15659.

Chugh, D., & Kern, M. C. (2016). A dynamic and cyclical model of bounded ethicality. *Research in Organizational Behaviour*, *36*, 85-100.

Clifford, S., Jewell, R. M., & Waggoner, P. D. (2015). Are samples drawn from Mechanical Turk valid for research on political ideology?. *Research & Politics*, 2(4).

Cialdini, R. B., Trost, M. R., & Newsom, J. T. (1995). Preference for consistency: The development of a valid measure and the discovery of surprising behavioral implications. *Journal of Personality and Social Psychology*, 69(2), 318.

Cohen, T. R., Wolf, S. T., Panter, A. T., & Insko, C. A. (2011). Introducing the GASP scale: a new measure of guilt and shame proneness. *Journal of personality and social psychology*, *100*(5), 947.

Cohn, A., Fehr, E., & Maréchal, M. A. (2014). Business culture and dishonesty in the banking industry. *Nature*, *516*(7529), 86-89.

Cojoc, D., & Stoian, A. (2014). Dishonesty and charitable behavior. *Experimental Economics*, 17(4), 717-732.

Conway, P., & Peetz, J. (2012). When does feeling moral actually make you a better person? Conceptual abstraction moderates whether past moral deeds motivate consistency or compensatory behavior. *Personality and Social Psychology Bulletin*, 38(7), 907-919.

Dana, J., Weber, R. A., & Kuang, J. X. (2007). Exploiting moral wiggle room: experiments demonstrating an illusory preference for fairness. *Economic Theory*, *33*(1), 67-80.

Dolan, P., & Galizzi, M. M. (2015). Like ripples on a pond: Behavioral spillovers and their implications for research and policy. *Journal of Economic Psychology*, 47, 1-16.

Effron, D. A., Bryan, C. J., & Murnighan, J. K. (2015). Cheating at the end to avoid regret. *Journal of personality and social psychology*, *109*(3), 395.

Effron, D. A., Cameron, J. S., & Monin, B. (2009). Endorsing Obama licenses favoring whites. *Journal of experimental social psychology*, *45*(3), 590-593.

Effron, D. A., Miller, D. T., & Monin, B. (2012). Inventing racist roads not taken: The licensing effect of immoral counterfactual behaviours. *Journal of personality and social psychology*, *103*(6), 916.

Effron, D. A., Monin, B., & Miller, D. T. (2013). The unhealthy road not taken: Licensing indulgence by exaggerating counterfactual sins. *Journal of Experimental Social Psychology*, *49*(3), 573-578.

Freedman, J. L., & Fraser, S. C. (1966). Compliance without pressure: the foot-in-the-door technique. *Journal of personality and social psychology*, *4*(2), 195.

Fischbacher, U., & Föllmi-Heusi, F. (2013). Lies in disguise—an experimental study on cheating. *Journal of the European Economic Association*, *11*(3), 525-547.

Gino, F., Ayal, S., & Ariely, D. (2013). Self-serving altruism? The lure of unethical actions that benefit others. *Journal of economic behaviour & organization*, *93*, 285-292.

Gino, F., & Bazerman, M. H. (2009). When misconduct goes unnoticed: The acceptability of gradual erosion in others' unethical behaviour. Journal of experimental Social psychology, 45(4), 708-719.

Gino, F., & Desai, S. D. (2012). Memory lane and morality: How childhood memories promote prosocial behaviour. Journal of personality and social psychology, 102(4), 743.

Gino, F., Norton, M. I., & Ariely, D. (2010). The counterfeit self: The deceptive costs of faking it. *Psychological science*, 21(5), 712-720.

Gino, F., Schweitzer, M. E., Mead, N. L., & Ariely, D. (2011). Unable to resist temptation: How self-control depletion promotes unethical behaviour. Organizational Behaviour and Human Decision Processes, 115(2), 191-203.

Gneezy, U. (2005). Deception: The role of consequences. The American Economic Review, 95(1), 384-

394.

Gneezy, A., Imas, A., Brown, A., Nelson, L. D., & Norton, M. I. (2012). Paying to be nice: Consistency and costly prosocial behaviour. *Management Science*, *58*(1), 179-187.

Gneezy, U., Imas, A., & Madarász, K. (2014). Conscience accounting: Emotion dynamics and social behaviour. *Management Science*, *60*(11), 2645-2658.

Gneezy, U., Kajackaite, A., & Sobel, J. (2017). Lying Aversion and the Size of the Lie. *American Economic Review*.

Gneezy, U., & List, J. (2014). *The why axis: hidden motives and the undiscovered economics of everyday life*. Random House.

Gneezy, U., Rockenbach, B., & Serra-Garcia, M. (2013). Measuring lying aversion. Journal of Economic Behaviour & Organization, 93, 293-300.

Gneezy, U., & Rustichini, A. (2000a). A fine is a price. *The Journal of Legal Studies*, *29*(1), 1-17.

Gneezy, U., & Rustichini, A. (2000b). Pay enough or don't pay at all. *The Quarterly Journal of Economics*, *115*(3), 791-810.

Greene, J. D., & Paxton, J. M. (2009). Patterns of neural activity associated with honest and dishonest moral decisions. *Proceedings of the National Academy of Sciences*, *106*(30), 12506-12511.

Hauser, O., Kraft-Todd, G., Rand, D., Nowak, M., & Norton, M. I. (2016, January). Invisible Inequality Leads to Punishing the Poor and Rewarding the Rich. In *Academy of Management Proceedings* (Vol. 1, p. 13841).

Hilbig, B. E., & Thielmann, I. (2017). Does everyone have a price? On the role of payoff magnitude for ethical decision making. Cognition, 163, 15-25.

Hollander, E. P. (1958). Conformity, status, and idiosyncrasy credit. *Psychological review*, *65*(2), 117.

Horton, J. J., & Chilton, L. B. (2010, June). The labor economics of paid crowdsourcing. In *Proceedings of the 11th ACM conference on Electronic commerce* (pp. 209-218). ACM.

Horton, J. J., Rand, D. G., & Zeckhauser, R. J. (2011). The online laboratory: Conducting experiments in a real labor market. *Experimental Economics*, 14(3), 399-425.

Jiang, T. (2013). Cheating in mind games: The subtlety of rules matters. *Journal of Economic Behaviour & Organization*, *93*, 328-336.

Jordan, A.H., & Monin, B. (2008). From sucker to saint: Moralization in response to self-threat. *Psychological Science*, 19(8), 683-689.

Jordan, J., Mullen, E., & Murnighan, J. K. (2011). Striving for the moral self: The effects of recalling past moral actions on future moral behaviour. *Personality and Social Psychology Bulletin*, *37*(5), 701-713.

Kajackaite, A., & Gneezy, U. (2017). Incentives and cheating. *Games and Economic Behaviour*, *102*, 433-444.

Kajackaite, A., & Gneezy, U. (2015). Lying costs and incentives. *UC San Diego Discussion Paper*.
Kessler, J. B., & Milkman, K. L. (2016). Identity in Charitable Giving. *Management Science*.

Kouchaki, M., & Gino, F. (2016). Memories of unethical actions become obfuscated over time. *Proceedings of the National Academy of Sciences*, *113*(22), 6166-6171.

Landers, R. N., & Behrend, T. S. (2015). An inconvenient truth: Arbitrary distinctions between organizational, Mechanical Turk, and other convenience samples. *Industrial and Organizational Psychology*, 8(2), 142-164.

Mazar, N., Amir, O., & Ariely, D. (2008). The dishonesty of honest people: A theory of self-concept maintenance. *Journal of marketing research*, *45*(6), 633-644.

Merritt, A. C., Effron, D. A., Fein, S., Savitsky, K. K., Tuller, D. M., & Monin, B. (2012). The strategic pursuit of moral credentials. *Journal of Experimental Social Psychology*, *48*(3), 774-777.

Merritt, A. C., Effron, D. A., & Monin, B. (2010). Moral self-licensing: When being good frees us to be bad. Social and personality psychology compass, 4(5), 344-357.

Miller, D. T., & Effron, D. A. (2010). Chapter three-psychological license: When it is needed and how it functions. Advances in experimental social psychology, 43, 115-155.

Monin, B., & Miller, D. T. (2001). Moral credentials and the expression of prejudice. *Journal of personality and social psychology*, *81*(1), 33.

Potters, J., & Stoop, J. (2016). Do cheaters in the lab also cheat in the field?. *European Economic Review*, *87*, 26-33.

Raihani, N. J., Mace, R., & Lamba, S. (2013). The effect of $1, $5 and $10 stakes in an online dictator game. *PloS one*, *8*(8), e73131.

Rudman, L. A. (2004). Social justice in our minds, homes, and society: The nature, causes, and consequences of implicit bias. *Social Justice Research*, *17*(2), 129-142.

Sezer, O., Gino, F., & Bazerman, M. H. (2015). Ethical blind spots: Explaining unintentional unethical behaviour. *Current Opinion in Psychology*, *6*, 77-81.

Shalvi, S., & De Dreu, C. K. (2014). Oxytocin promotes group-serving dishonesty. *Proceedings of the National Academy of Sciences*, *111*(15), 5503-5507.

Shalvi, S., Eldar, O., & Bereby-Meyer, Y. (2012). Honesty requires time (and lack of justifications). *Psychological science*, *23*(10), 1264-1270.

Shalvi, S., Handgraaf, M. J., & De Dreu, C. K. (2011). Ethical manoeuvring: Why people avoid both major and minor lies. *British Journal of Management*, *22*(s1).

Shu, L. L., Gino, F., & Bazerman, M. H. (2011). Dishonest deed, clear conscience: When cheating leads to moral disengagement and motivated forgetting. *Personality and Social Psychology Bulletin*, *37*(3), 330-349.

Starmer, C., & Sugden, R. (1991). Does the Random-Lottery Incentive System Elicit True Preferences? An Experimental Investigation. *American Economic Review , 81* (4), 971-978.

Truelove, H. B., Carrico, A. R., Weber, E. U., Raimi, K. T., & Vandenbergh, M. P. (2014). Positive and negative spillover of pro-environmental behavior: An integrative review and theoretical framework. *Global Environmental Change*, 29, 127-138.

Von Hippel, W., & Trivers, R. (2011). The evolution and psychology of self-deception. *Behavioural and Brain Sciences*, *34*(01), 1-16.

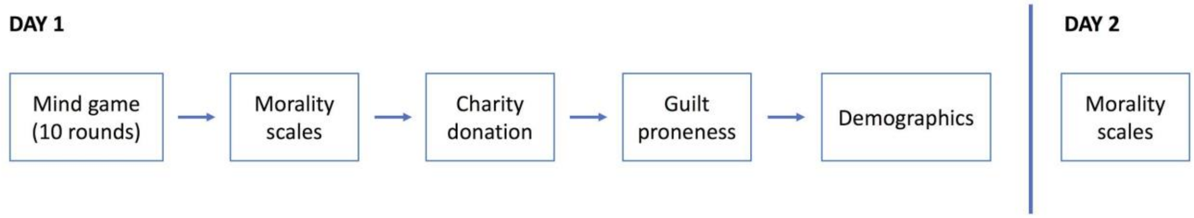
Zhang, T., Fletcher, P. O., Gino, F., & Bazerman, M. H. (2015). Reducing bounded ethicality: How to help individuals notice and avoid unethical behaviour.

Zhong, C. B., & Liljenquist, K. (2006). Washing away your sins: Threatened morality and physical cleansing. *Science*, *313*(5792), 1451-1452.

Zhong, C. B., Liljenquist, K. A., & Cain, D. M. (2009). Moral self-regulation. *Psychological perspectives on ethical behaviour and decision making*, 75-89.

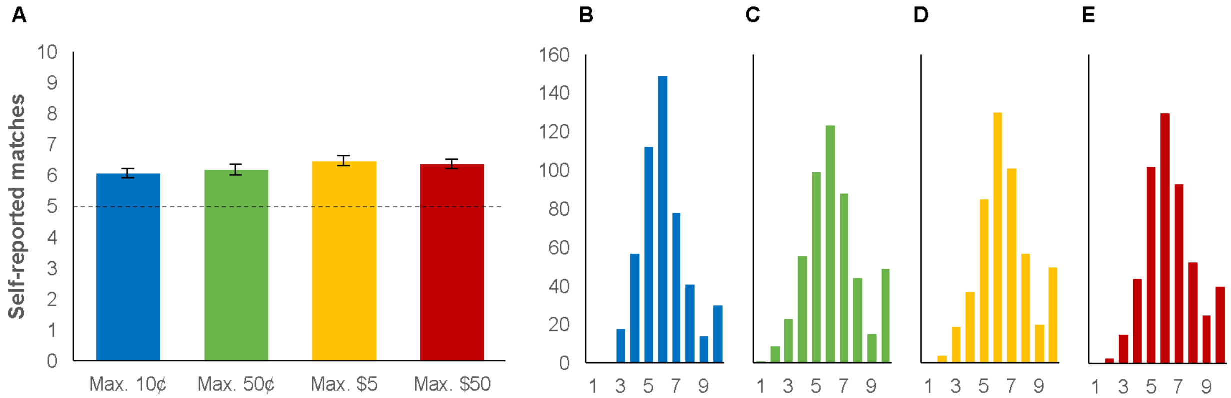
Acknowledgements. We are grateful to a grant from Google that funded this research.

Figure 1. Illustration of each stage of the two-day study.



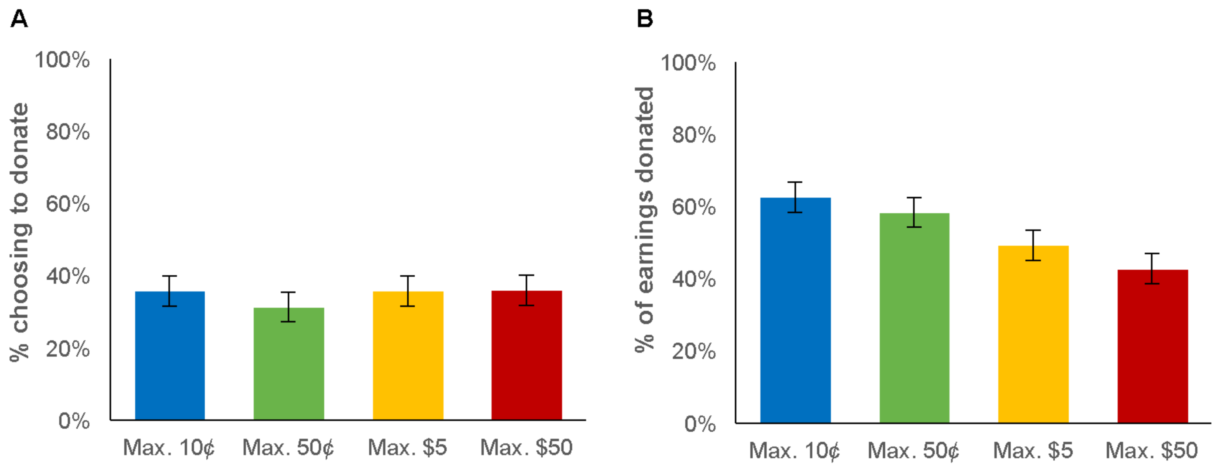
Note. Participants played the repeated mind game for 10 rounds, followed by scales that elicited their self-reported morality. Then participants were asked if they wanted to give some of their earnings to a charity and, if so, what percentage of their earnings. Participants then filled out a number of exploratory items (including the 'negative behavioural evaluation' subscale of the Guilt and Shame Proneness (GASP) scales which we refer to as "guilt proneness" scale) and demographics. All participants received an invitation the next day to participate in a short follow-up survey that included the same morality scales from the day before.

Figure 2. Dishonest behaviour changed little as incentives increased.



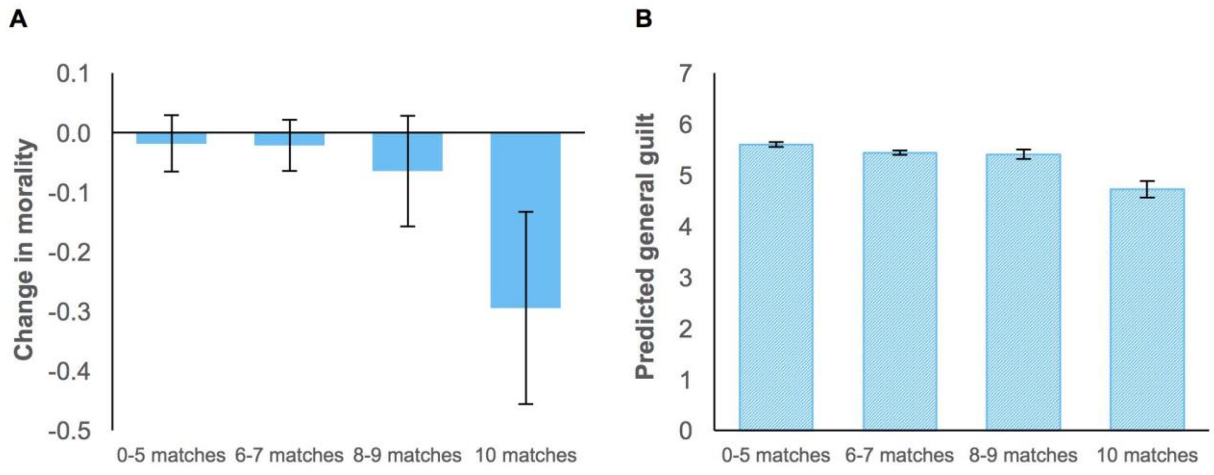
Note. *(A)* As incentives increase, participants report significantly more matches in the 10 rounds of the coin flipping task. However, the effect size is significant but small: the average number of matches increases by less than 5% as incentives multiply by a factor of 500. *(B-E)* The distributions of matches reported across 10 rounds by increasing stake sizes (maximum payoffs: (B) \$0.10, (C) \$0.50, (D) \$5.00, or (E) \$50.00) look qualitatively similar.

Figure 3. Higher stakes did not affect the likelihood to engage in charitable behaviour but decreased the percentage of earnings donated.



Note. (A) The likelihood of making a donation does not change with stake size. *(B)* However, participants who choose to donate are affected by the size of their potential earnings: as stake increases donors choose to give away a smaller share of their earnings to a charity of their choice. This effect is driven by self-reported morality: participants who report feeling more moral in the high-stakes condition donate less to charity.

Figure 4. Participants who cheated maximally mispredicted their perceived morality.



Note. (A) Participants who are most likely dishonest (reporting 10 matches out of 10 coin flips) report feeling less moral one day after the coin flipping task. In contrast, participants who do not cheat (0-5 matches), cheat only a little (6-7 matches), or cheat moderately (8-9 matches) report similar moral feelings the next day. *(B)* However, participants who cheat the most are bad at predicting their own feelings in the future: the most dishonest participants say they would feel less guilty after committing wrongdoing than all other participants, but their predictions do not match actual changes in reduced self-reported morality.

Table 1. Donation levels—the percentage of earnings donated—are lower in higher-stakes conditions.

VARIABLES	(1)	(2)	(3)
Condition: Max. $0.10	(omitted)	(omitted)	(omitted)
Condition: Max. $0.50	-4.236	-3.810	-4.266
	(3.296)	(3.222)	(12.186)
Condition: Max. $5.00	-13.403***	-12.116***	-19.994
	(3.132)	(3.096)	(12.346)
Condition: Max. $50.00	-19.850***	-18.931***	-28.730*
	(2.861)	(2.825)	(12.236)
# Matches Reported		-3.300***	-4.082**
		(0.665)	(1.486)
Max. $0.50 X # Matches Reported			0.094
			(1.974)
Max. $5.00 X # Matches Reported			1.327
			(1.967)
Max. $50.00 X # Matches Reported			1.654
			(1.961)
Constant	62.570***	81.634***	86.148***
	(2.305)	(4.531)	(9.084)
Observations	700	700	700
R-squared	0.072	0.105	0.107

Note. Linear regression predicting percentage of earnings donated by condition dummies and self-reported number of matches in the coin-flipping task (for those participants who decided to donate). Robust standard errors in parentheses. *** p<0.001, ** p<0.01, * p<0.05.

Table 2. Self-reported morality does not change with stake size, but is predicted by the number of matches reported in the coin-flipping task.

VARIABLES	(1)	(2)	(3)
Condition: Max. $0.10	(omitted)	(omitted)	(omitted)
Condition: Max. $0.50	-0.071	-0.062	0.093
	(0.046)	(0.045)	(0.163)
Condition: Max. $5.00	0.028	0.059	0.230
	(0.046)	(0.046)	(0.171)
Condition: Max. $50.00	-0.016	0.006	0.259
	(0.046)	(0.045)	(0.172)
# Matches Reported		-0.076***	-0.052**
		(0.009)	(0.019)
Max. $0.50 X # Matches Reported			-0.025
			(0.026)
Max. $5.00 X # Matches Reported			-0.028
			(0.026)
Max. $50.00 X # Matches Reported			-0.041
			(0.027)
Constant	4.178***	4.642***	4.493***
	(0.033)	(0.063)	(0.123)
Observations	2,015	2,015	2,015
R-squared	0.002	0.037	0.038

Note. Linear regression predicting self-reports of morality by condition dummies and self-reported number of matches in the coin-flipping task. Robust standard errors in parentheses. *** p<0.001, ** p<0.01, * p<0.05.

Table 3. In the highest stake condition, self-reporting feeling more moral *reduces* the percentage of earnings that donors are willing to give to charity.

VARIABLES	(1) Max. $0.10	(2) Max. $0.50	(3) Max. $5	(4) Max. $50	(5) All	(6) Interaction	(7) Interaction
Morality Index	5.414	-0.629	-1.753	-6.425**	-0.851	5.414	4.335
	(3.408)	(3.152)	(3.695)	(2.439)	(1.581)	(3.107)	(3.058)
Condition: Max. $0.50					-4.420	21.843	27.471
					(3.079)	(18.434)	(18.134)
Condition: Max. $5.00					-13.396***	17.704	17.799
					(2.964)	(20.952)	(20.574)
Condition: Max. $50.00					-19.899***	31.138	29.294
					(2.958)	(18.943)	(18.604)
Max. $0.50 X Morality						-6.043	-7.358
						(4.291)	(4.221)
Max. $5.00 X Morality						-7.168	-6.882
						(4.777)	(4.691)
Max. $50.00 X Morality						-11.839**	-11.202**
						(4.342)	(4.265)
# Matches Reported							-3.406***
							(0.659)
Constant	39.082**	60.925***	56.786***	70.220***	66.264***	39.082**	63.441***
	(14.959)	(13.207)	(16.198)	(10.570)	(7.174)	(13.640)	(14.199)
Observations	179	159	180	182	700	700	700
R-squared	0.014	0.000	0.001	0.037	0.073	0.083	0.117

Note. Linear regression predicting percentage of earnings donated by self-reported morality (separately by condition, Cols. 1-4) and condition dummies (Cols. 5-6). The number of self-reported matches in the coin-flipping task is controlled for in Col. 7. Robust standard errors in parentheses. *** p<0.001, ** p<0.01, * p<0.05.

Appendix

Table A1. Linear regression using stake size (Cols. 1-2) and the natural logarithm of stake size (Cols. 3-4) as predictors of the average number of matches reported.

VARIABLES	(1)	(2)	(3)	(4)
Stake size	0.000	0.000		
	(0.000)	(0.000)		
ln(stake size)			0.055**	0.054**
			(0.017)	(0.018)
Male		-0.169		-0.168
		(0.085)		(0.084)
MTurk experience (years)		0.036		0.035
		(0.026)		(0.026)
Familiarity with coin flips		-0.070		-0.074
		(0.089)		(0.089)
Materialism index		0.018		0.017
		(0.029)		(0.028)
Social esteem index		-0.034		-0.034
		(0.027)		(0.025)
Competitiveness index		-0.020		-0.029
		(0.031)		(0.030)
Altruism index		0.046		0.045
		(0.028)		(0.028)
Donation frequency dummies		Yes		Yes
Age group dummies		Yes		Yes
Education dummies		Yes		Yes
U.S. region dummies		Yes		Yes
Constant	6.236***	6.962***	6.118***	6.881***
	(0.048)	(0.668)	(0.064)	(0.668)
Observations	2,015	2,015	2,015	2,015
R-squared	0.001	0.031	0.005	0.035

Note. Robust standard errors in parentheses. *** p<0.001, ** p<0.01, * p<0.05.

Table A2. The propensity to donate to a charity does not vary by condition.

VARIABLES	(1)	(2)	(3)
Condition: Max. $0.10	(omitted)	(omitted)	(omitted)
Condition: Max. $0.50	-0.199	-0.189	-0.580
	(0.134)	(0.135)	(0.488)
Condition: Max. $5.00	-0.001	0.059	-0.092
	(0.132)	(0.133)	(0.505)
Condition: Max. $50.00	0.010	0.054	-0.027
	(0.132)	(0.133)	(0.509)
# Matches Reported		-0.150***	-0.179**
		(0.027)	(0.060)
Max. $0.50 X # Matches Reported			0.065
			(0.078)
Max. $5.00 X # Matches Reported			0.026
			(0.079)
Max. $50.00 X # Matches Reported			0.014
			(0.081)
Constant	-0.584***	0.320	0.489
	(0.093)	(0.187)	(0.366)
Observations	2,015	2,015	2,015

Note. Logit regression predicting choosing to donate some of one's earnings to a charity. Robust standard errors in parentheses. *** p<0.001, ** p<0.01, * p<0.05.

Table A3. Linear regression of self-reported morality immediately after the coin-flipping task predicted by group (based on number of matches reported) and stake size condition.

VARIABLES	(1) Max. $0.10	(2) Max. $0.50	(3) Max. $5	(4) Max. $50	(5) All	(6) All
Group 1	(omitted)	(omitted)	(omitted)	(omitted)	(omitted)	(omitted)
Group 2	-0.044	-0.041	-0.016	0.071	-0.003	-0.044
	(0.068)	(0.075)	(0.070)	(0.075)	(0.036)	(0.070)
Group 3	0.029	-0.050	-0.147	-0.190	-0.096	0.029
	(0.106)	(0.111)	(0.093)	(0.100)	(0.051)	(0.108)
Group 4	-0.588***	-0.748***	-0.607***	-0.832***	-0.693***	-0.588***
	(0.136)	(0.119)	(0.108)	(0.129)	(0.061)	(0.139)
Max. $0.10					(omitted)	(omitted)
Max. $0.50						-0.028
						(0.073)
Max. $5.00						0.066
						(0.078)
Max. $50.00						-0.004
						(0.076)
Group 2 X $0.50						0.003
						(0.099)
Group 2 X $5.00						0.028
						(0.102)
Group 2 X $50.00						0.115
						(0.101)
Group 3 X $0.50						-0.079
						(0.151)
Group 3 X $5.00						-0.176
						(0.147)
Group 3 X $50.00						-0.219
						(0.146)
Group 4 X $0.50						-0.160
						(0.179)
Group 4 X $5.00						-0.018
						(0.181)
Group 4 X $50.00						-0.243
						(0.187)
Constant	4.230***	4.202***	4.297***	4.226***	4.235***	4.230***
	(0.050)	(0.054)	(0.055)	(0.057)	(0.027)	(0.052)
Observations	500	507	503	505	2,015	2,015
R-squared	0.039	0.077	0.069	0.100	0.068	0.075

Note. Groups 1, 2, 3, and 4 are made up of participants who reported 0-5 matches, 6-7 matches, 8-9 matches, and 10 matches, respectively. We refer to Group 1 as "mostly honest" while Group 4 are the "maximal cheaters." Robust standard errors in parentheses. *** p<0.001, ** p<0.01, * p<0.05.

Table A4. Linear regression of difference between self-reported morality immediately after the coin-flipping task and one day later predicted by group (based on number of matches reported) and stake size condition.

VARIABLES	(1) Max. $0.10	(2) Max. $0.50	(3) Max. $5	(4) Max. $50	(5) All	(6) All
Group 1	(omitted)	(omitted)	(omitted)	(omitted)	(omitted)	(omitted)
Group 2	-0.031	-0.059	0.054	0.035	-0.003	-0.031
	(0.062)	(0.062)	(0.080)	(0.073)	(0.034)	(0.068)
Group 3	-0.049	0.007	-0.106	-0.001	-0.047	-0.049
	(0.091)	(0.096)	(0.104)	(0.095)	(0.048)	(0.100)
Group 4	-0.222	-0.100	-0.370**	-0.350**	-0.276***	-0.222
	(0.124)	(0.106)	(0.122)	(0.117)	(0.058)	(0.136)
Max. $0.10					(omitted)	(omitted)
Max. $0.50						0.041
						(0.069)
Max. $5.00						-0.034
						(0.076)
Max. $50.00						-0.032
						(0.072)
Group 2 X $0.50						-0.028
						(0.095)
Group 2 X $5.00						0.085
						(0.100)
Group 2 X $50.00						0.065
						(0.096)
Group 3 X $0.50						0.055
						(0.143)
Group 3 X $5.00						-0.058
						(0.139)
Group 3 X $50.00						0.047
						(0.134)
Group 4 X $0.50						0.122
						(0.177)
Group 4 X $5.00						-0.148
						(0.177)
Group 4 X $50.00						-0.128
						(0.175)
Constant	-0.015	0.025	-0.049	-0.047	-0.018	-0.015
	(0.045)	(0.045)	(0.062)	(0.056)	(0.026)	(0.050)
Observations	345	356	335	377	1,413	1,413
R-squared	0.009	0.005	0.042	0.030	0.018	0.026

Note. Groups 1, 2, 3, and 4 are made up of participants who reported 0-5 matches, 6-7 matches, 8-9 matches, and 10 matches, respectively. We refer to Group 1 as "mostly honest" while Group 4 are the "maximal cheaters." Robust standard errors in parentheses. *** p<0.001, ** p<0.01, * p<0.05.

Table A5. Linear regression predicting guilt proneness by group (based on number of matches reported).

VARIABLES	(1) All	(2) Max. 10c	(3) Max. 50c	(4) Max. $5	(5) Max. $50
Group 1	(omitted)	(omitted)	(omitted)	(omitted)	(omitted)
Group 2	-0.161*	-0.291*	-0.062	-0.141	-0.138
	(0.068)	(0.128)	(0.139)	(0.145)	(0.133)
Group 3	-0.191*	-0.334	-0.297	-0.065	-0.139
	(0.096)	(0.199)	(0.206)	(0.193)	(0.178)
Group 4	-0.878***	-0.686**	-0.884***	-0.946***	-0.900***
	(0.115)	(0.255)	(0.222)	(0.224)	(0.229)
Constant	5.600***	5.602***	5.568***	5.526***	5.700***
	(0.051)	(0.095)	(0.101)	(0.113)	(0.101)
Observations	2,015	500	507	503	505
R-squared	0.028	0.020	0.034	0.037	0.030

Note. Groups 1, 2, 3, and 4 are made up of participants who reported 0-5 matches, 6-7 matches, 8-9 matches, and 10 matches, respectively. We refer to Group 1 as "mostly honest" while Group 4 are the "maximal cheaters." Robust standard errors in parentheses. *** p<0.001, ** p<0.01, * p<0.05.

Table A6. Average reported coin toss matches by stake size condition.

CONDITION	Average reported matches
Max. $0.10	6.08
Max. $0.50	6.19
Max. $5.00	6.48
Max. $50.00	6.37
Overall	6.28
Observations	2,015

Table A7. Average donation rates and average donation proportion by stake size condition.

CONDITION	Average donation rates	Average donation proportion
Max. $0.10	35.80%	62.57%
Max. $0.50	31.36%	58.33%
Max. $5.00	35.79%	49.17%
Max. $50.00	36.04%	42.72%
Overall	34.74%	53.00%
Observations	2,015	700

Note. Average donation proportion only for those participants who decided to donate.

Table A8. Self-reported morality immediately after and one day after the task, by stake size condition.

CONDITION	Immediately after	One day after
Max. $0.10	4.204	4.157
Max. $0.50	4.101	4.094
Max. $5.00	4.235	4.156
Max. $50.00	4.155	4.091
Overall	4.172	4.123
Observations	1,413	1,413

Table A9. Self-reported morality immediately after and one day after the task, by group (based on number of matches reported).

CONDITION	Immediately after	One day after
Group 1	4.242	4.224
Group 2	4.237	4.216
Group 3	4.189	4.124
Group 4	3.521	3.227
Overall	4.172	4.123
Observations	1,413	1,413

Note. Groups 1, 2, 3, and 4 are made up of participants who reported 0-5 matches, 6-7 matches, 8-9 matches, and 10 matches, respectively. We refer to Group 1 as "mostly honest" while Group 4 are the "maximal cheaters."

Table A10. Guilt proneness by group (based on number of matches reported).

CONDITION	Guilt proneness
Group 1	5.600
Group 2	5.439
Group 3	5.409
Group 4	4.722
Overall	5.430
Observations	2,015

Note. Groups 1, 2, 3, and 4 are made up of participants who reported 0-5 matches, 6-7 matches, 8-9 matches, and 10 matches, respectively. We refer to Group 1 as "mostly honest" while Group 4 are the "maximal cheaters."