scientific reports

Check for updates

OPEN

Dishonest behavior can transition to continuous ethical transgressions

Crystal Reeck[1]✉ & Dan Ariely[2]

Ethical decision-making research often focuses on singular events. However, in our daily lives we are regularly tempted to behave dishonestly, and little is known about whether and how repeated ethical deviations can worsen over time. Building upon research on dishonesty and moral disengagement, we test the effect of repeated transgressions on dishonesty. Across six experiments we provide evidence that repeated transgressions can lead to a transition to continuous dishonesty. Individual dishonest acts are thus not independent events, but rather can compound and perpetuate pervasive unethical behavior. We find that making one's self-interested actions more salient hinders the transition to continuous cheating, while unfair financial deprivation, self-serving rewards, and gradual change -- factors that previous research has shown to facilitate moral disengagement – all facilitate the transition to continuous cheating. Moreover, we show that moral disengagement mediates the effect of gradual change on the transition to continuous dishonesty.

Keywords Dishonesty, Moral disengagement, Morality, Ethical decision making

"*The last initial deviation from the truth is multiplied later a thousandfold.*"

Aristotle[1].

Introduction

The marketplace and the workplace are characterized by abundant opportunities to behave dishonestly. For example, employees may engage in dishonest acts, from taking credit for others' work to stealing office supplies for personal use to violating professional codes of conduct to overstating expenses or hours billed to clients. Individuals are often victims of others' immoral transgressions as well, from doctors prescribing medicines produced by pharmaceutical companies they receive benefits from to financial advisers steering investors towards options with high commissions without disclosure of their conflicts of interest[2]. Despite the fact that honesty is a universal value and many of us have strong beliefs about our own morality[3], it is easy to deviate occasionally from the path of honesty. However, each individual lapse may not occur independently, and previous ethical transgressions may compound to make future transgressions more likely. In the present paper, we consider whether these infractions are momentary lapses of ethicality from which people quickly recover or if acts of dishonesty can beget subsequent acts of dishonesty. Specifically, after repeated transgressions the nature of behavior might transition so that dishonest behavior becomes a pervasive and persistent pattern. After the first few exaggerations about one's expenses, does one at some point start inflating all subsequent potential deductions? Do we gingerly overbill a couple hours, only to at some point find ourselves overbilling unabashedly?

If our transgressions were just isolated bumps on an otherwise honest path, the cumulative effect of our dishonest acts would simply equal their individual magnitudes. But, if at some point unethical behavior becomes a pervasive pattern, then the cost of our transgressions could be much higher, since each additional transgression would also increase the likelihood of eschewing honest behavior altogether. That is, people may engage in continuous cheating, where they only provide dishonest responses. Thus, it is important to know, whether such a transition to continuous cheating can actually happen, and if so, what are some factors that facilitate or hinder it.

Moral disengagement & transitions to continuous cheating

Bandura[4,5] demonstrated that individuals can selectively engage in self-regulatory processes such as reconstruing agency (through for example, diffusion of responsibility) or reconstructing morality (through, for example, justifications or rationalizations[6] of ethical misconduct) to reduce self-censure and thereby clearing the way for unethical behaviors. That is, through a selective, intrapsychic and motivated cognitive reasoning process,

[1]Fox School of Business, Temple University, 1801 Liacouras Walk, Philadelphia, PA 19122, USA. [2]Fuqua School of Business, Duke University, 1 Towerview Drive, Durham, NC 27708, USA. ✉email: crystal.reeck@temple.edu

individuals can shift their moral judgment, which allows common people to temporarily behave unethically [We focus on Bandura's original process-conceptualization of moral disengagement (i.e. a state-like construct) rather than a propensity-conceptualization (i.e. trait-like construct).]. This dissonance-reducing process has been coined moral disengagement[7–9].

Bandura later[10] went on to discuss gradual disengagement of internalized moral standards and moral control over repeated enactments. Though his focus was on how average considerate individuals transform into cruel ones that engage in inhumane practices, the idea he argued for is that as self-reproof diminishes, the level or severity of transgression increases (see also[11,12]) until eventually acts originally regarded as unacceptable and violating ones' internal standards can be performed with little personal anguish, guilt, or self-sanctions[6]. That is, extreme practices become thoughtlessly routinized.

Morality is a domain in which the study of an escalation of behavior seems central but unlike other contexts of undesirable behaviors such as binge eating and various kinds of addictions[13,14], it has not received much empirical attention. Anecdotal evidence surrounding the scandals involving companies such as Theranos, Enron, Lehman Brothers, and Wells Fargo or longitudinal studies about criminal offences suggest that many people do not start out 'bad' but rather that their sporadic transgressions worsen over time[15].

Yet, in the domain of dishonesty, most empirical investigations focus on singular events (e.g.,[16]) or two subsequent events such as in the literature on moral compensation, licensing, and consistency[17–20]. Such research is important to understanding the basic mechanism underlying relatively isolated decisions to be honest or dishonest, but it neglects more complex, possible path dependencies whereby the accumulation of past dishonest acts influences subsequent dishonesty.

In addition, while previous research has focused primarily on factors affecting individual differences in the propensity to morally disengage (see Newman et al., 2020 for a systematic review), to date little is known about more immediate context-specific factors that facilitate or hinder moral disengagement. In fact, in their recent review, Newman et al. (2020) call for more research to examine the situational factors that lead individuals to morally disengage over time. To date some of the few factors that have been suggested to hinder moral disengagement processes is to make one's self-interested actions and their consequences more salient[21]or to get individuals to be more prevention focused (Welsh et al. 2015). Factors that have been suggested to facilitate moral disengagement processes include resource depletion (e.g.,[22–25]), perceptions of injustice[26,27], self-serving rewards[28], and gradual as opposed to abrupt changes (Welsh et al. 2015).

Previous research on a series of opportunities to be dishonest

Particularly relevant to our focus on the context of a series of opportunities to be dishonest is research on the cheat-at-the-end effect[29] and the slippery slope effect[11,30,31]. Effron and colleagues (2015) had individuals face a finite series of opportunities to cheat and examined to what extent knowing whether they had arrived at the last opportunity to cheat changed their likelihood to engage in self-serving dishonesty in that last instance. The authors found that individuals were more likely to cheat and their magnitude of cheating was higher when they thought they were facing the last opportunity. Furthermore, the authors found that this cheat-at-the-end effect was more likely driven by a desire to avoid feeling regret rather than alternative mechanisms based on moral self-licensing, ego-depletion, or a slippery slope. However, the authors suggest that their documented effect may be limited to relatively short series with fewer than 20 cheating opportunities. Thus, these findings do not address how transitions to continuous dishonesty may emerge across a longer series of choices or in circumstances that do not include the last known opportunity to act unethically.

Changing the perspective to the dishonest actors, Welsh and colleagues (2015) and Garrett and colleauges (2016) examined whether actors engaged in a slippery slope of increasingly unethical behavior. Those latter two contributions are most closely related to our current research focus as they elucidated how unethical behavior unfolds over time.

Welsh and colleagues (2015) showed that as the *potential* for unethicality increased gradually over time rather than abruptly (e.g., by gradually versus abruptly increasing the payoff resulting from cheating or the opportunities to cheat), the magnitude of cheating when facing the last opportunity or set of opportunities to do so, was highest in the gradual change condition. Note, this was true despite the fact that the incentives to cheat in the final period were identical across the gradual-change and the abrupt-change conditions. Furthermore, the authors found that this slippery-slope effect was driven by moral disengagement. Specifically, Welsh et al. (2015) adapted several items from Moore et al.[32] to assess the degree of rationalizations that participants engaged in to moralize their actions in the task. This state-level measure of moral disengagement happened immediately before the final set of opportunities to cheat and revealed that gradual changes in payoffs increased rationalizations, which in turn resulted in more cheating. Finally, Welsh and colleagues (2015) provided evidence that a prevention focus intervention mitigated the slippery slope effect and reduced dishonesty in the final period. [Prevention focus was induced through a word completion task, using words such as *secure*, *vigilance*, *prevention* and giving participants instructions (a) to write down an outcome that they strongly wanted to avoid and (b) to describe the strategies they planned to use to avoid this outcome.] However, this research did not examine whether cheating eventually transitioned to become continuous.

Unlike Welsh and colleagues (2015), Garrett and colleagues (2016) did not study the effects of a series of increasing opportunities to cheat on dishonest behavior but rather the effects of a series of random opportunities to cheat on dishonest behavior. In addition, they used functional MRI to examine the underlying mechanism. In essence, Garrett and colleagues (2016) found that the magnitude of cheating increased over a series of opportunities to cheat. Furthermore, they found evidence of signal reduction in the amygdala (a brain region implicated in emotional processing) over past dishonest behaviors, consistent with adaptation, that is, Bandura's idea of gradual moral disengagement. However, same as Welsh and colleagues (2015), this research did not establish that people ever behave maximally dishonestly, transitioning to cheating constantly.

Overview of studies

Building on this body of work in behavioral ethics research, we consider the very extreme end-result of facing a series of opportunities to cheat over time. That is, the present work moves past this prior research by seeking to explore the existence of a possible transition to uncontrolled unethicality, that is continuous maximal cheating. Such a transition to continuous cheating may arise as a result of Bandura's gradual moral disengagement, such that repeated opportunities to cheat at some point no longer trigger much personal anguish and therefore, no longer hinder immoral actions. Understanding the existence of a transition to continuous cheating and whether some factors can slow down or prevent this transition is of collective importance, as people's unethical actions can have negative impacts on other people. At the same time, there are reasons to anticipate that such effects may not emerge in the moral domain, given the widespread desire to maintain a positive moral-self concept[3]. We also seek to contribute to the existing literature by identifying choice architecture interventions to curb the transition to continuous maximal cheating.

The present experiments examine the phenomenon of a transition to continuous cheating in a task featuring repeated incentives to cheat. Across six incentive-compatible laboratory experiments, we demonstrate that a subset of people eventually transition to continuous cheating. Pilot-Experiment 1 establishes the existence of a transition to continuous cheating in a repeated decision-making task. Experiment 2 replicates this finding and tests a decision point intervention that offers a costly opportunity to remove incentives for dishonesty as a means to encourage subsequent honesty and rehabilitate those who have transitioned to continuous cheating. It also allows for an in-depth examination of how people who transition to continuous cheating compare to others in their subsequent judgments, decisions, and actions. Experiment 3 yet again replicates the existence of a transition to continuous cheating effect and, in addition, tests for the role of making one's self-interested actions more salient (see[21]) for evidence that this hinders moral disengagement) in counteracting the continuous cheating. Finally, in Experiments 4a-4c, we reanalyzed data from published research that used repeated opportunities to cheat and find evidence supporting the phenomenon of a transition to continuous cheating in addition to revealing that financial deprivation/unfairness, self-serving rewards, and gradual change -- factors that previous research has shown to facilitate moral disengagement -- all increase the transition to continuous cheating. Moreover, in Experiment 4c we find that moral disengagement mediates the effect of gradual change on the transition to continuous dishonesty. The materials, data files, and accompanying analysis codes for studies 1–3 are available on ResearchBox at https://researchbox.org/443&PEER_REVIEW_passcode=XHNUZP.

Pilot-Experiment 1: existence of a transition to continuous cheating

Methods

We sought to determine whether repeated opportunities for self-serving dishonesty could lead to transitions to continuous cheating. In our pilot-study (first outlined in[33]), thirty-six students from a North-American University (gender and age information not collected) participated in this within-subjects design in exchange for $5 plus their earnings in the task. They provided informed consent, and all procedures were approved by the Princeton University Institutional Review Board. All methods were carried out in accordance with relevant guidelines and regulations.

The study used a computer-based visual perception ("Dots") task to study deception (for other research that has used the task for that purpose, see e.g.,[25,34–36]. The task consisted of two identical rounds of 100 trials. Each trial displayed a pattern of 20 dots scattered inside a box divided by a diagonal line. The dots were displayed for one second after which participants were asked to identify whether the left or right side of the diagonal line had more dots. Each round consisted of 52 trials with more dots on the left, 34 trials with more dots on the right, and 14 ambiguous trials with an equal number of dots on both sides.

Upon completion of the first 100 trials, participants learned that this was the first round and that they would repeat the same experience, only this time for pay. Participants were told that because most people could more easily estimate the number of dots on the left than on the right side they would get 0.5 cents for each trial identified as having more dots on the left side and 5 cents for each trial identified as having more dots on the right side[37]. Since this procedure paid participants based on their response and not their accuracy, it created a conflict of interest on trials with more dots on the left: on these trials participants faced a dilemma between following the task instructions (identifying which side had more dots) by truthfully indicating what they saw (earning only 0.5 cents) or dishonestly indicating that there were more dots on the right (earning 10 times the amount) [In a pre-test of the instructions for the Dots-Task's for pay-round published by Mazar and Hawkins[38] individuals ($N = 29$) unanimously agreed that the task was to correctly identify which side had more dots, and going for higher pay instead of accuracy was rated as very dishonest.].

Results and discussion

The results showed that in the for-pay round, participants more often identified trials as having more dots on the higher-pay (i.e. right) side (overall $M = 70.9$, SEM = 3.2) than they did in the round for no pay ($M = 41.4$, SEM = 0.8, $t(35) = 8.83$, $p < .001$, $d = 1.47$). This behavioral bias toward the higher-pay side consisted of three components (see Table 1 for details):

(i) Participants were more accurate on trials that had more dots on the higher-pay (i.e. right) side ($t(35) = 5.70$, $p < .001$, $d = 0.95$; note that participants were already fairly accurate to begin with, correctly identifying 90% of trials in the no-pay round);
(ii) Participants interpreted ambiguous trials (trials that had the same amount of dots on either side) more often in favor of the higher pay side ($t(35) = 9.97$, $p < .001$, $d = 1.66$; note that participants answered roughly at random in the no-pay round, identifying 45% as having more dots on the higher-pay side, but then switched to self-benefiting, biased responses 81% of the time in the for-pay round);

Actual Trial Type (# of trials)	Mean % (SEM)	
	Practice Round (i.e. no pay)	For-Pay Round
More dots on the higher pay side (right; # of trials = 34)	90.0 (1.1)	96.7 (0.8)
Ambiguous (# of trials = 14)	44.6 (2.4)	80.8 (3.4)
More dots on the lower pay side (left; # of trials = 52)	8.7 (0.9)	51.3 (5.2)

Table 1. Percent of trials by actual trial type identified by respondents as having more Dots on the right (i.e., higher pay) side in experiment 1.

Actual Trial Type (# of trials)	Before Continuous Cheating	During Continuous Cheating	Mean Comparison
More dots on the higher pay side (right; # of trials = 34)	0.72 (0.09)	0.40 (0.02)	t(18) = 3.86, p = .00 d = 0.89
Ambiguous (# of trials = 14)	0.68 (0.10)	0.36 (0.02)	t(18) = 3.50, p = .003 d = 0.80
More dots on the lower pay side (left; # of trials = 52)	0.66 (0.07)	0.39 (0.03)	t(17) = 3.89, p = .001 d = 0.92

Table 2. Response times in the for-pay round by trial type in experiment 1. Response times are in seconds with SEM in parentheses.

(iii) Participants were more likely to lie for higher pay on trials that had more dots on the lower pay (i.e. left) side ($t(35) = 8.10$, $p < .001$, $d = 1.35$; note that participants' error rate was fairly low in the no-pay round with only 9% of trials misclassified, but then multiplied roughly six times once we introduced the possibility for self-benefiting dishonesty).

More importantly, we found evidence for a transition to continuous cheating. Here and throughout the entire paper we defined "transitioning to continuous cheating" as providing the higher pay response without interruption on the final 10% of trials that pose a dilemma of dishonesty (trials with more dots on the lower-paying side) as well as all intervening trials (those with more dots on the higher-paying side and, if available, ambiguous trials). [The final 10% of trials that featured more dots on the lower-paying side might not necessarily only occur in the final 10% of all trials (due to ambiguous trials and/or trials featuring more dots on the higher-pay side). The choice of using the final 10% of trials was arbitrary and was kept constant across all studies in this paper, including those where we re-analyzed others' published work (Experiments 4a-c).] By this definition, more than half of all participants, 52.8% ($n = 19$) eventually transitioned to continuous cheating in the for-pay round of this experiment. There was wide variation among our participants in the onset of continuous cheating, with the earliest transition occurring on the third trial and the latest transition occurring on the 89th trial ($M = 41.2$, SD = 27.7).

We anticipated that the transition to continuous cheating would be accompanied by a disengagement from moral judgment, as participants would no longer be making a decision on each trial but rather would be merely providing the more profitable response each time. An analysis of participants' response times supported this assertion (Table 2). In the round for pay, response times decreased significantly across all three trial types once participants switched from sporadic cheating to continuous cheating (trials with more dots on the left: $t(17) = 3.89$, $p = .001$; ambiguous trials: $t(18) = 3.50$, $p = .003$; trials with more dots on the right: $t(18) = 3.86$, $p = .001$). This pattern of results implies that participants no longer deliberated on the correct response and instead simply always provided the more profitable response, consistent with moral disengagement.

Experiment 2: an exploration into subsequent judgments, decisions, and behaviors after transitioning to continuous cheating

Experiment 1 provided evidence that a transition to continuous dishonesty is not just a theoretical possibility but empirically observable. In addition, we observed a reduction in reaction times after transitions to continuous cheating. This latter finding implies that after a transition to continuous cheating participants no longer deliberated on the correct response and instead merely provided the more profitable response each time.

Experiment 2 was designed to not only replicate the findings from Experiment 1, but also to explore how a transition to continuous cheating may affect subsequent judgments, decisions, and actions. Additionally, in other domains of repetitive behavior such as eating or spending, adding a decision point can curb behavior[39]. In this context, adding a decision point may draw participants' attention back to the conflict of interest and their behavior, which might curb moral disengagement. Toward that end, we added a decision point-intervention after the for-pay round, followed by a second for-pay round of the dots task, and a question to test participants' accurate perception of the dots task.

Methods

Seventy-eight students from a North-American University (50 females; age information not collected) participated in this within-subjects design in exchange for $5 plus their earnings in the task. They provided informed consent, and all procedures were approved by the University of Toronto Institutional Review Board. All methods were carried out in accordance with relevant guidelines and regulations.

The procedure used a version of the dots task similar to that from Experiment 1 but, since the focus of our study was on cheating, without ambiguous trials (i.e. there were 60 trials with more dots on the left, and 40 trials with more dots on the right). Furthermore, unlike Experiment 1 participants earned more for left responses (5.0 cents) as opposed to for right responses (0.5 cents). Importantly, we added a second for-pay round in the dots task. Specifically, after having experienced same as in Experiment 1 two identical rounds of 100 trials—the first as practice without any pay and the second for pay—we surprised our participants by announcing that we would "repeat the exact same task for real pay," but that this time they could choose the payment scheme. In particular, participants could choose to retain the unequal payment scheme from the previous round (which would expose them to conflicts of interests) or choose to switch to one paying an equal amount of 2.5 cents for either side, thus, allowing participants to focus on identifying the side with more dots without any opportunities (i.e. dilemma) for self-benefiting dishonesty. This intervention to change the incentive structure of the task acted as a decision point and could have made participants' experienced conflicts of interests and their behaviors salient to them. After the final round, participants were asked to estimate how many of the 100 trials actually had more dots on the higher pay side (i.e. left).

As each of the rounds included 60 trials with more dots on the higher pay (left) and 40 on the lower pay (right) side, the unequal payment scheme represented the payoff-maximizing option regardless of whether individuals cared about honesty or not. That is, a person who always answered honestly could make $3.20 and a person who always selected the highest paying option whether it was the honest thing to do or not would make $5. In comparison, a person, who chose the equal payment scheme would make exactly $2.50. Consequently, if one knew that 60% of trials had more dots on the higher pay side (or at least assumed a 50:50 split) and calculated the potential payoffs, choosing the equal payment scheme only made sense if participants wanted to avoid incentives to be dishonest and were willing to sacrifice earnings for that purpose. This decision point therefore could offer an opportunity to engage in costly precommitment. However, choosing the equal payment scheme may instead simply be a result of depleted resources, if participants viewed it as the less cognitively taxing option, or of a desire to avoid further uncomfortable situations (i.e. facing conflicts of interests).

Results and discussion

First For-Pay Round. Behavior. As in Experiment 1, participants provided more profitable responses during the first round for pay ($M = 68.5$, SEM = 1.6) than the initial no-pay, practice round ($M = 63.5$, SEM = 1.2, $t(77) = 5.00$, $p < .001$, $d = 0.57$). Importantly, we also found evidence for a transition to continuous cheating. Specifically, 19.2% of participants ($n = 15$) transitioned to continuous cheating. There was substantial variation in the trial on which people transitioned to continuous cheating (Minimum = 1st trial, Maximum = 87th trial, $M = 37.9$, SD = 38.9). We also examined whether continuous cheating was associated with gender, but found no significant relationship ($p = .712$).

As examined in Experiment 1, during the first round for pay participants' response times decreased after transitioning to continuous cheating (before continuous cheating: $M = 0.59$s, SEM = 0.04s versus during continuous cheating $M = 0.42$s, SEM = 0.02s; $t(8) = 5.31$, $p = .001$, $d = 1.77$), suggesting that they no longer deliberated on the correct response and instead defaulted to providing the profitable response.

Decision Point Intervention. After completion of the first round for pay and learning about the existence of a second round for pay, 56.4% of participants ($n = 44$) chose the suboptimal equal payment scheme for that last round of 100 trials. Of note, the post-experiment survey revealed that participants' perceptions of the number of trials with more dots on the higher pay side (left) were fairly accurate. Their mean estimate was not significantly different from the correct response of 60 trials ($M = 59.8$, $SD = 13.2$, $t(77) = 0.16$, $p = .871$). Thus, it is possible that participants realized (if they engaged in some form of deliberation) that choosing equal pay was a costly decision. In sum, more than half of participants chose to remove incentives for self-benefitting dishonesty.

To examine whether behavior in the first for-pay round was related to the subsequent payment scheme choice, we segmented participants based on their behavior in the first for-pay round. Towards this end, since Experiment 2 did not have ambiguous trials, we could calculate a biased error rate ([percent of trials wrongfully identified as having more dots on the left, higher pay side] - [percent of trials wrongfully identified as having more dots on the right, lower pay side]) for each participant and segment participants into three groups based on their degree of self-benefiting responses [Of note, since an unbiased error rate does not preclude inaccurate responding (e.g., one could randomly respond and thus, be unbiased, but as a consequence also have very low accuracy), we confirmed our segmentation by analyzing the percentage of trials for which participants provided the accurate answer. As one would expect, overall, accuracy during the first round for pay was significantly lower for those classified as continuous cheaters ($M = 65.5\%$, $SD = 8.3\%$) compared to those either classified as fairly dishonest ($M = 80.8\%$, $SD = 7.9\%$, $t(29) = 5.30$, $p < .001$, $d = 1.91$) or fairly honest individuals ($M = 92.1\%$, $SD = 2.4\%$, $t(60) = 12.36$, $p < .001$, $d = 5.91$).]:

1. Continuous Cheaters ($n = 15$),
2. Fairly Dishonest individuals (i.e., 'biased error rate' >10% of trials) without evidence for continuous cheating ($n = 16$), and.
3. Fairly Honest individuals (i.e., 'biased error rate' <=10% of trials) without evidence for continuous cheating ($n = 47$).

Indeed, the number of participants choosing to avoid further opportunities for self-benefiting dishonesty differed significantly between the three participant segments ($\chi^2(2) = 6.90$, $p = .032$). The equal payment scheme was selected by 68.1% of those who had been fairly honest versus only 43.8% of those who had been fairly dishonest ($\chi^2(1) = 3.00$, $p = .083$), and 33.3% of Continuous Cheaters (in comparison to fairly honest: $\chi^2(1) = 5.71$, $p = .017$). The difference between the Continuous Cheaters and Fairly Dishonest individuals was not significant ($\chi^2(1) = 0.35$, $p = .552$).

We followed up with additional analyses in the first round for pay of the Continuous Cheaters segment to explore potential differences between those who subsequently opted to continue versus eliminate further opportunities for self-benefitting dishonesty, but couldn't find any significant differentiators. First, these two subgroups did not transition to continuous cheating at different time points, $t(13) = 0.78$, $p = .449$. [Note, however, that this test may be underpowered due to the smaller number of participants in this segment. The effect size for this comparison is $d = 0.43$, suggesting that even if a difference does exist that we were unable to detect here, it is not a large effect.] Second, they did not differ with respect to either overall accuracy, $t(13) = 0.30$, $p = .769$, or their biased error rates, $t(13) = 0.63$, $p = .538$.

Second For-Pay Round. Behavior. Participants' payment scheme choice for the final for-pay round had an effect on their subsequent behavior ($F(1,72) = 81.58$, $p < .001$, $n_p^2 = 0.53$). First, in terms of biased error rate: those choosing to switch to equal pay decreased their biased error rate in comparison to the preceding round for pay ($M = -39.6\%$, SEM = 4.2%), while those choosing to continue with unequal pay increased their biased error rate in comparison to the preceding round for pay ($M = +11.1\%$, SEM = 3.7%). In fact, the biased error rate in this final round was not significantly different from 0 for those who chose to switch to equal pay ($t(43) = 0.59$, $p = .559$), meaning that those who switched to equal pay were not biased toward the higher pay side when giving their responses. This decrease in the biased error rate from the first to the second round for pay was largest for those who had previously transitioned to continuous cheating (interaction; $F(2,72) = 22.58$, $p < .001$, $n_p^2 = 0.39$; see Fig. 1). Second, inconsistent with a resource depletion account, the accuracy rates of those who chose to continue with equal pay were very high (M = 88.0%, SD = 11.5%) and not significantly different ($t(42) = 1.45$, $p = .219$) between those who had transitioned to continuous cheating in the previous round for-pay (percent of trials with correct responses: $M = 73.0\%$, $SD = 26.0\%$) and those who had not (percent of trials with correct response among fairly honest and fairly dishonest participants combined: $M = 90.0\%$, SD = 6.8%).

Post-Experiment Survey Accuracy Judgment. At the end of the dots task, when participants were asked to estimate how many of the 100 trials truly had more dots on the higher-paying side, there was no difference in

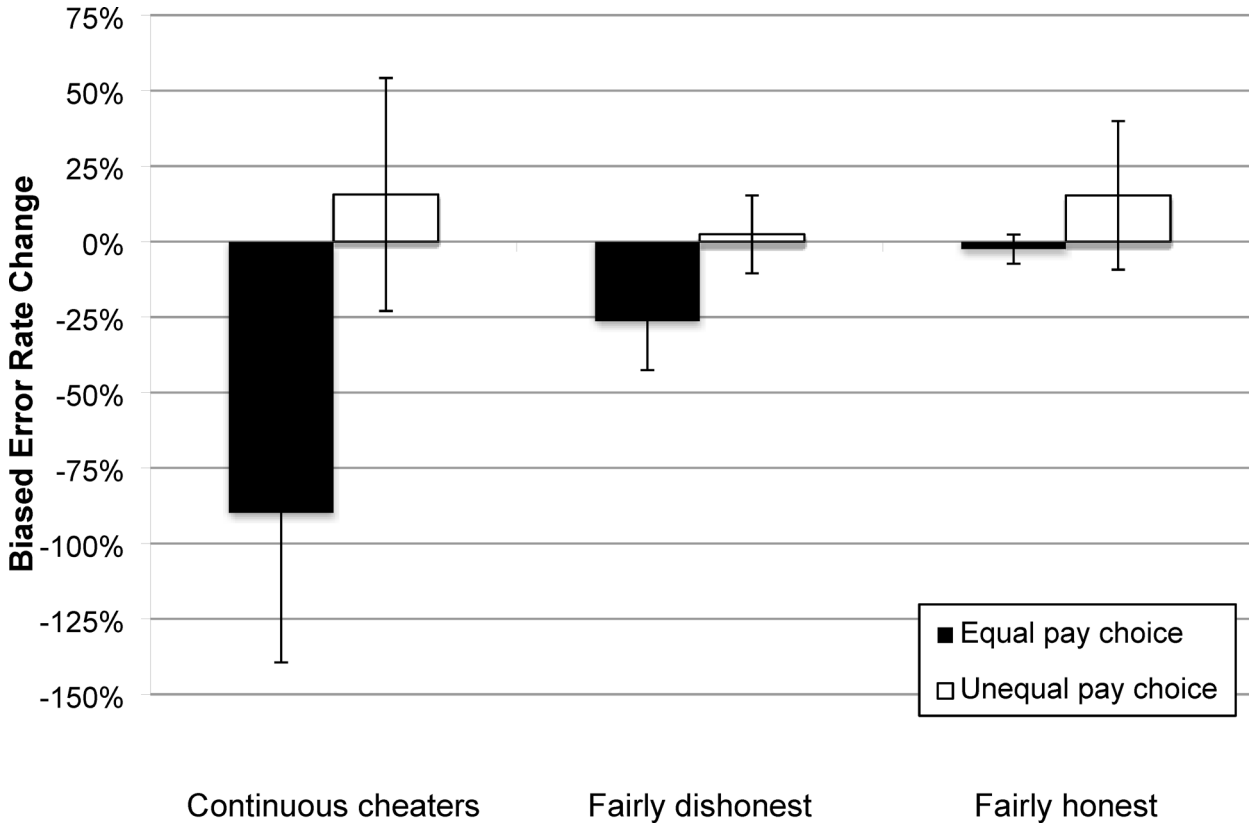

Fig. 1. 'Biased error rate' change from the for-pay round before to the for-pay round after the choice of the payment scheme in Experiment 2. Decreases indicate that participants were less dishonest in their responses in the final round after the choice of payment scheme. Error bars represent standard deviations. Total $N = 78$. Continuous cheaters $n = 15$; Fairly dishonest $n = 16$; Fairly honest $n = 47$.

estimates between those who had been classified as Continuous Cheaters ($M = 54.6$, $SD = 10.5$) and those who had not been classified as such, ($M = 61.0$, $SD = 13.6$), $t(76) = 1.70$, $p = .093$. In addition, there was no statistically significant difference between the estimates of participants who chose equal pay ($M = 61.6$, $SD = 11.8$) and unequal pay ($M = 57.4$, $SD = 14.7$; $t(76) = 1.40$, $p = .165$). Overall, given that after three rounds of the dots task and the fact that participants were not paid for accuracy the mean estimate was not significantly different from the correct response of 60 trials ($M = 59.8$, $SD = 13.2$, $t(77) = 0.16$, $p = .871$), it is less likely that those who transitioned to continuous cheating or those that choose unequal pay did so due to resource depletion.

Together, the findings in this experiment point toward two insights. First, inserting a decision point, a simple interruption that offers participants a choice in terms of how to proceed (an explicit decision-making opportunity), may be a promising simple intervention to get individuals to pause and think about their (dis) honest actions, helping some of them (even those that had transitioned to continuous cheating) to snap out of a disengaged mode of decision making (see related findings about partitioning and decision points in the context of consumption activities by[39]). Second, because accuracy in the dots-task for those who selected equal pay for the final for-pay round was fairly high and did not differ by behavior in the preceding round, it suggests that participants, including those who had transitioned to continuous cheating, were not resource depleted or demotivated. The same conclusion can be drawn from the fairly accurate estimates in the post-experiment survey.

Experiment 3: saliency mitigates the transition to continuous cheating

Methods

Experiments 1 and 2 established the existence of a transition to continuous cheating. In Experiment 3, we sought to examine the role of saliency in hindering the transition to continuous cheating. In particular, previous research on ethical decision making has shown that moral disengagement (i.e. reconstruing agency and morality) is hindered when one's self-interested unethical actions and their consequences are made salient[21]. In line with those previous findings, in Experiment 3, we manipulated the information participants viewed about their behavior during the task to make emerging biases in their responses more or less salient during task performance.

Additionally, we sought to examine how those who transitioned to continuous cheating viewed their behavior compared to those who did not transition to continuous cheating. As a result, we added a questionnaire to the end of the task measuring their perceptions of their own behavior, their judgments of continuous cheating, and their reactions to an ethical transgression in another domain.

Methods

Seventy-nine participants were recruited from a North American University community, provided informed consent, and all procedures were approved by the Duke University Institutional Review Board. All methods were carried out in accordance with relevant guidelines and regulations. The procedure used in Experiment 3 was the same as the one in Experiment 2 in that there were no ambiguous trials (unlike Experiment 2 there was an equal split of trials with more dots on the left and on the right) and that the left side was associated with higher pay (5 cents vs. 0.5 cents for each trial identified as having more dots on the right side), and that there were a total of three rounds (one practice round followed by two rounds for pay). Unlike Experiment 1 and 2, each trial featured 100 dots rather than 20 dots, with a vertical line dividing the display. In addition, there was one practice round and one for pay round and the order of the 100 trials in each of these two rounds was randomized for each participant. Finally, and most importantly, the information participants viewed between trials was manipulated between participants. Half of the participants ($N = 40$) viewed on each trial their total earnings in the task thus far (e.g., "Your TOTAL in cents so far: 149"), similar to the procedure in Experiments 1 and 2. For the other half of the participants ($N = 39$), however, their total responses for each side were instead displayed on each trial (e.g., "Your TOTAL responses so far: Left: 27 Right: 28"). This latter feedback was intended to make one's actual responses more salient to participants and, potentially, mitigate the likelihood of biased responses toward the higher pay side.

In this experiment, we also sought to examine whether participants who transitioned to continuous cheating viewed their behavior as problematic. Specifically, we wanted to examine whether they perceived their performance as less accurate and whether they viewed their responses as inappropriate. Towards that end, after completing the dots task participants completed a short survey (see the Methodological Details Appendix for all questions).

Results and discussion

As in Experiments 1 and 2, participants were significantly more biased toward the left (i.e. higher pay) side in the for-pay round (68.8%) than the practice round (57.3%; i.e. without pay), $t(78) = 8.34$, $p < .001$, $d = 0.94$. Accuracy for trials with more dots on the "Left" (i.e. the higher pay side in the for-pay round) versus those with more dots are the "Right" (i.e. the lower pay side in the for-pay round) was higher for both the practice-round, $t(78) = 5.09$, $p < .001$, $d = 0.57$, and the for-pay round, $t(78) = 8.85$, $p < .001$, $d = 1.00$, indicating that the increased incentive for left responses did bias participants' responding. This difference in accuracy was significantly smaller in the practice round than the for-pay round, $F(1, 78) = 69.80$, $p < .001$, $n_p^2 = 0.47$, likely due to the fact that responses in the for-pay round actually impacted participants' compensation.

We again conservatively defined "switching to continuous cheating" as providing the higher pay response ("Left") without interruption on the final 10% of trials that pose a dilemma of dishonesty (trials with more dots on the lower-paying side) as well as all intervening trials (those with more dots on the higher-paying side). Critically, the feedback manipulation significantly impacted the extent to which participants transitioned to continuous cheating, $\chi^2(1) = 11.86$, $p = .001$, as participants who viewed their actual responses (i.e. # of Left and

Right response) were much less likely (10.3%) than those who simply viewed their earnings (45.0%) to exhibit the effect. Even for those participants who did transition to continuous cheating, those who viewed their actual responses made significantly less money, $t(20) = 2.56$, $p = .019$, $d = 1.42$, and transitioned to continuous cheating marginally later, $t(20) = 1.81$, $p = .086$, $d = 1.00$, than those who viewed their earnings. [Once again, there was substantial variation in the trial on which people transitioned to continuous cheating (Minimum = 1st trial, Maximum = 90th trial, $M = 28.7$, SD = 33.3).] Thus, making one's actual behaviors (in this case the responses people gave) salient during task performance appeared to counteract the transition to continuous cheating. By making one's behavior more apparent (through feedback about one's distribution of responses).

After completing the task and paying themselves, participants completed a short survey to gauge their perceptions of the task and their performance on it. There were no differences between the two conditions in how accurate, acceptable, appropriate, and ethical participants viewed their performance on the task (p's > 0.1). We also asked them to imagine they had transitioned to continuous cheating (we asked them to imagine they started the task with equal accuracy on both sides, but that by the end they were only accurate on the higher-pay side), and asked participants to judge how acceptable exhibiting this behavior would be and found no differences (p's > 0.2). Participants in the two conditions also did not differ with respect to how they generally viewed themselves (p's > 0.6) or how they judged a different type of ethical transgression (stealing copier paper from work, p's > 0.1). Thus, the manipulation did not alter overall judgments of the task or change general judgments of morality, suggesting a more specific influence on moral disengagement during task performance.

Finally, we analyzed the perceptions of those who had transitioned to continuous cheating in the for-pay round ($n = 22$) compared with those who had not. ($n = 57$) and found that they rated their behavior as significantly less accurate ($t(76) = 2.40$, $p = .024$, $d = 0.75$) compared to participants who did not transition to continuous cheating. That is, participants who had transitioned to continuous cheating were aware that their accuracy was poorer and, more importantly, viewed their behavior as less ethical ($t(76) = 3.63$, $p = .001$, $d = 0.93$). [Accuracy was objectively poorer among participants transitioning to continuous cheating (55.6% accurate) compared to other participants (81.2% accurate), $t(77) = 12.59$, $p < .001$.] Critically, the two groups made similar estimates of the number of trials that featured more dots on the higher pay ("Left") side during the for-pay round ($p = .363$), indicating that continuous cheating did not emerge from believing that providing the left response was more likely to be accurate during task completion.

Of note, those who had transitioned to continuous cheating and those who had not transition did not differ in their judgments of continuous cheating when it was described (p's > 0.05), indicating that both groups viewed continuous cheating in general as similarly inappropriate. Finally, these two groups also did not differ in their judgment of a different type of ethical transgression (stealing office supplies from work; all p's > 0.1) except for how ashamed they thought they would feel. Those that had transitioned to continuous cheating indicated they would feel significantly less ashamed if they took a ream of paper ($t(76) = 2.34$, $p = .022$, $d = 0.60$) than those who had not transitioned to continuous cheating. This later finding is consistent with moral disengagement.

Experiments 4a-4c: moderators of the transition to continuous cheating

Methods

We next sought to establish the existence of the phenomenon of a transition to continuous cheating and other potential moderators of it by re-analyzing previously published data by other research teams that examined a series of ethical decisions. This approach is particularly powerful as these studies were originally collected with other goals in mind. If we can detect evidence of transitions to continuous cheating and identify conditions that make such transitions more likely, these findings will give additional credence to our own observations in Studies 1–3. These studies each manipulated a factor that has been hypothesized to increase moral disengagement: perceptions of unfairness[26,27], self-serving rewards[28], and gradual as opposed to abrupt changes (Welsh et al. 2015). In each case, we re-analyze data that focused only on individual actions of dishonesty to examine whether continuous cheating occurred. Moreover, in Experiment 4c we test moral disengagement as a potential mediator predicting the transition to continuous cheating.

Experiment 4a: ("Unfair") financial deprivation facilitates the transition to continuous cheating

Study 4a re-analyzed data from Sharma et al.[36] that investigated how financial deprivation altered ethical decision making. In their Experiment 1, participants ($N = 89$) engaged in four rounds for pay of the Dots task but each round was preceded by a slot machine game, which either earned or cost them $2.50 each round. [Their Dots task was the same as the one we used in our Experiment 1, except that Sharma et al. (2014) had no ambiguous trials (i.e., they had 60 trials with more dots on the left and 40 trials with more dots on the right).] Unbeknown to the participants, the slot machine was rigged and participants were randomly assigned to a condition in which they always won $2.50 (financially non-deprived participants) or always lost $2.50 (financially deprived participants). For financially deprived participants, the repeated losses may have felt unfair, which prior research has shown can increase moral disengagement[26,27]. We therefore predicted more transitions to continuous cheating among the financially deprived participants.

First, we found evidence for a transition to continuous cheating (i.e. participants cheated continuously on the last 10% of trials that posed a dilemma of dishonesty). Second, and more importantly, we found (a) significantly more participants transitioning to continuous cheating in the financially deprived condition than in the non-deprived condition (repeated-measures ANOVA between-subjects effect: $F(1, 87) = 10.96$, $p = .001$), (b) the number of participants transitioning to continuous cheating significantly increased over the four rounds (repeated-measures ANOVA within-subjects effect: $F(3, 85) = 7.21$, $p < .001$) and, (c) the increase was significantly higher for the financially-deprived participants compared to the non-deprived participants (repeated-measures ANOVA interaction: $F(3, 85) = 2.71$, $p = .05$; see Table 3). Finally, focusing only on those who transitioned to

	Non-Deprived $N=42$	Deprived $N=47$
Round 1	19.0% ($N=8$)	38.3% ($N=18$)
Round 2	26.2% ($N=11$)	53.2% ($N=25$)
Round 3	23.8% ($N=10$)	55.3% ($N=26$)
Round 4	26.2% ($N=11$)	66.0% ($N=31$)

Table 3. Percentage of participants transitioning to continuous cheating by round and condition in experiment 4a, re-analysis of Sharma et al., 2014.

	Non-Deprived	Deprived
Round 1	37.5 (36.2)	35.6 (32.0)
Round 2	22.5 (32.0)	20.6 (29.0)
Round 3	2.5 (4.7)	7.3 (16.6)
Round 4	16.1 (33.6)	3.0 (7.4)

Table 4. Onset of continuous cheating by round and condition in experiment 4a, reanalysis of Sharma et al., 2014. Mean trial number with standard deviations in parentheses.

continuous cheating, the onset of their continuous cheating was not statistically different in the first three rounds (round 1: $t(24)=0.14$, $p=.89$; round 2: $t(34)=0.18$, $p=.86$; round 3: $t(34)=-0.89$, $p=.38$), but was significantly earlier (by 13 trials) among the financially deprived than the financially non-deprived participants in the final round ($t(40)=2.08$, $p=.04$; see Table 4).

Experiment 4b: personal incentives facilitate the transition to continuous cheating

Study 4b re-examined data from Garrett et al. (2016). In this experiment, participants ($N=25$ in an fMRI scanner) were in an advisor role offering estimates of the number of pennies in jars to others who would rely on their estimates. The experiment manipulated within-subjects (in separate randomized blocks) whether participants were incentivized to provide higher estimates which would (1) increase their own payment but harm the advisee (self-serving, other-harming), (2) increase both participants' payouts (self-serving, other-serving), or (3) increase only the advisee's payout while decreasing their own payout (self-harming, other-serving). Prior research has demonstrated that self-serving rewards can increase moral disengagement[28], so we anticipated more transitions to continuous cheating in the self-serving conditions compared to the self-harming condition.

First, we found evidence for a transition to continuous cheating (i.e. participants cheated continuously on the last 10% of trials that posed a dilemma of dishonesty). Second, and more importantly, we found differences in the percent of participants transitioning to continuous cheating between conditions: 52.0% of participants did so in the self-serving, other-harming block, and 72.0% of participants did so in the self-serving, other-serving block. Conversely, only 8.0% of participants transitioned to continuous cheating in the self-harming, other-serving block. The lower rate of transitions to continuous cheating observed in this later block is expected, given that cheating is disincentivized for both parties involved (advisor and advisee).

The rate of the transition to continuous cheating was significantly higher in the self-serving, other-serving session ($p<.001$) and the self-serving, other-harming session ($p=.003$) compared to the self-harming, other-serving session. While there were numerically more participants who exhibited a transition to continuous cheating in the self-serving, other-serving session compared to the self-serving, other-harming session, this difference was not statistically significant ($p=.180$). [We note that this was a within-subjects design, so the same participants exhibited the transition to continuous cheating in some blocks and not in others.]

Experiment 4c: gradual changes facilitates the transition to continuous cheating, mediated by moral disengagement

Study 4c re-examined data from Studies 3 and 4 from Welsh et al.[11]. Since the findings were the same in both studies, we detail our results from the authors' Study 3 ($N=92$) here and refer for the details from their Study 4 ($N=207$) to the supplement.

Welsh et al.'s (2015) Study 3 included completing 50 rounds of a version of the dots task and looking at whether gradual changes made people more likely to provide self-interested responses than abrupt changes. Specifically, during the 50 rounds of the task participants were either gradually offered more and more opportunities to cheat as the experiment progressed or abruptly in the final five rounds given opportunities to cheat. These gradual changes facilitated moral disengagement, which enabled more cheating. However, they did not analyze their results to examine whether continuous cheating occurs and whether gradual compared to abrupt changes made continuous cheating more likely, as we would predict.

Again, first, we found evidence for the existence of a transition to continuous cheating (i.e. participants cheated continuously on the last 10% of trials that posed a dilemma of dishonesty). Second, and more importantly, we found differences in the percent of participants that transitioned to continuous cheating between conditions. Third, we found mediation by moral disengagement.

Overall, 16.3% of participants transitioned to continuous cheating. Importantly, those in the gradual change condition were significantly more likely to do so (26.1%) than those in the abrupt change condition (6.5%; $x^2(1) = 6.45$, $p = .011$). Additionally, this experiment measured moral disengagement, prior to the last 10% of trials. [Welsh et al. (2015) adapted four items from Moore et al. to specifically refer to moral disengagement related to the reprehensible behavior (i.e. cheating) in the dots task (Cronbach's alpha = 0.72). The authors focused primarily on the moral justification/rationalization mechanism, advantageous comparison mechanism, and displacement of responsibility mechanism: (1) "It's ok if someone selected the side with fewer dots as long as they had a good reason for doing so," (2) "Selecting the side with fewer dots is okay as long as you tried," (3) "Considering the ways that people grossly misrepresent themselves, it's not a big deal to occasionally select the wrong side," and (4) "People shouldn't be held accountable for not selecting the correct side with more dots given how quickly the dots disappeared."] That is, the authors interrupted the dots task just before its end, allowing us to examine whether moral disengagement mediates the observed increase in the percent of participants transitioning to continuous cheating among the gradual change group. We used a bootstrap mediation model[40] with 1,000 resamples, revealing indeed a significant indirect effect of gradual versus abrupt change on the transition to continuous cheating mediated by moral disengagement (indirect effect = 0.47, 95% CI [0.02, 1.22]). [Their last 10% of trials only consisted of trials that posed a dilemma for dishonesty. That means that continuously giving the higher pay response on at least those last 10% of trials represents "continuous cheating," according to our definition in this paper. Of note, all of the participants classified according to this definition as transitioning to continuous cheating had done so before the break (range of trial onset: 1 to 42).]

General discussion

The standard theory of crime and punishment assumes that individuals engage in an isolated, rational cost-benefit analysis whenever deciding whether to be honest or dishonest. This assumption leaves little room for the possibility of a slippery slope – a scenario where initially sporadic dishonest transgressions can worsen to a point where an individual entirely morally disengages and gives in to dishonest behavior.

The current paper documents people's vulnerability to committing repeated ethical transgressions and extends previous research on dishonesty. In particular, our findings suggest that dishonest acts can be super-additive over time and lead to an increase of unethical behavior to the point where it becomes pervasive and routine. The present findings are consistent with Bandura's[4,5] theory of gradual moral disengagement, such that repeated opportunities to cheat at some point no longer trigger much personal anguish and therefore, no longer hinder immoral actions. To our knowledge, the present research is the first empirical documentation of such a transition to continuous cheating.

In addition, we document, in part by re-analyzing previously published data examining a series of ethical decisions, that some of the factors that have been shown by previous research to facilitate or hinder moral disengagement to also affect a transition to continuous cheating: (1) salience of one's behavior (i.e. drawing attention to the accumulation of dishonest actions; Experiment 3), (2) financial well-being, (3) self-serving rewards, and (4) gradual as opposed to abrupt changes. We also demonstrate that moral disengagement does indeed mediate the transition to continuous cheating (Experiment 4c). Finally, we find evidence for a transition to continuous cheating across an array of circumstances. For example, we find the effect both when people are shown a running total of their earnings (Experiments 1, 2, 3, and 4a) and when they are not (Experiments 4b and 4c). The transition to continuous cheating emerges when another person might be harmed and using a task more akin to choices encountered in the real-world (Experiment 4b). Importantly, by detecting this effect in previously published data, we provide particularly compelling evidence that this effect emerges. As these data were originally collected with other goals in mind, our identification of the transition to continuous cheating lends additional credence to the existence of the effect.

At the same time, our findings reveal that those who transition to continuous cheating are not unaware of the ramifications of their actions (Experiment 3). They rightly view their own performance as less accurate compared to participants who do not transition to continuous cheating, and they perceive their behavior as less ethical. Yet, both those who had and those who had not transitioned to continuous cheating viewed a transition to continuous cheating in general as well as a different type of ethical transgression (stealing office supplies from work) as similarly unethical and inappropriate, indicating that continuous cheating likely did not emerge due to a redefinition of what honesty entails. Interestingly though, those who had transitioned to continuous cheating thought they would feel less ashamed from engaging in a hypothetical different ethical transgression. Note also that opting into an equal payment scheme – even at a wage premium – appears to promote honest behavior even among those who had transitioned to continuous cheating. However, as payment scheme was not randomly assigned here, strict causal claims are challenging.

The question arises whether there are individual differences in the onset of continuous cheating in the face of repeated opportunities to do so. Specifically, those who exhibit a transition to continuous cheating may exhibit different amounts of cheating before the transition and/or differ in their tendency to morally disengage (i.e. propensity-conceptualization of moral disengagement, see Schaefer and Bouwmeester (2020) or Moore, 2015). With regards to the former, we find mixed evidence. In Experiment 1, *before* switching their response pattern, the participants who eventually transitioned to continuous cheating displayed only limited cheating ($M = 29.3\%$) and were similarly dishonest as the participants who did not transition to continuous cheating ($M = 23.2\%$, $t(33) = 0.75$, $p = .460$). However, in Experiment 2 those who eventually transitioned to continuous cheating, *before* switching their response pattern displayed more cheating ($M = 46.8\%$) than the participants who did not

transition to continuous cheating ($M = 17.8\%$; $t(70) = 4.18$, $p < .001$). Future research could further explore this issue to establish whether onsets vary across individuals or are generally stable.

In the present research, we document the transition to continuous cheating following repeated opportunities for self-serving ethical transgressions. These findings represent a first step, in that we examined repeated incentives for transgressions of the same kind only and in close succession, such as when illegally downloading a multitude of songs. That is, the same task is performed by participants throughout our experiment. However, perhaps when the types of incentives and transgressions or the time between those repeated transgressions vary, they are less likely to result in a transition to continuous cheating as it may be easier for individuals to view them as single events. Relatedly, if transgressions occur in different tasks they may be perceived as isolated events and therefore are less likely to promote a transition to continuous dishonesty. One interesting aspect of our re-analyses of others' published data (Experiment 4b) is that they reveal that the transition to continuous cheating can emerge and disappear within-subjects across runs with different payment structures. Changing the payment structure may serve as a sort of reset or decision point, allowing people to re-engage with the task and altering their behavior. Future work could examine other interventions that help people return to ethical behavior after having transitioned to continuous cheating. Relatedly, our experiments used relatively small monetary stakes, in alignment with existing laboratory studies of cheating (Effron et al., 2015; Mazar et al., 2008; Shalvi et al., 2012; Welsh et al., 2015). Future work could examine how these processes might change at substantially higher stakes (Zak et al., 2022). Our sample also was not culturally diverse, and future work could examine how these processes differ across different cultures and backgrounds. Another interesting question pertains to differences in processing strategies before and after transitioning to continuous cheating. Our reaction time data from Experiments 1 and 2 seem to suggest that processing strategies do change. We posit that the observed speeding of response times following a transition to continuous cheating reflects disengagement from the task, but additional work could further explore this possibility.

In sum, our findings suggest that a situational transition to continuous cheating is not just a hypothetical construct and present various mediators of the phenomenon. Given that in our daily lives we face repeated opportunities for self-serving immoral actions, and that unethical conduct is very costly to societies[41], the implications of our and future related research may be substantial for policy aimed at curbing it.

Data availability

The materials, data files, and accompanying analysis codes for studies 1-3 are available on ResearchBox at https://researchbox.org/443&PEER_REVIEW_passcode=XHNUZP.

Received: 5 February 2025; Accepted: 1 July 2025
Published online: 08 July 2025

References

1. Aristotle *On the Heavens. Book I, Part 5*. 350 B.C.
2. Cain, D. M., Loewenstein, G. & Moore, D. A. When sunlight fails to disinfect: Understanding the perverse effects of disclosing conflicts of interest. *J. Consum. Res.* **37** (5), 836–857 (2011).
3. Schwartz, S. H. Are there universal aspects in the structure and contents of human-values. *J. Soc. Issues*. **50** (4), 19–45 (1994).
4. Bandura, A. & Pervin, L. A. Self-regulation of motivation and action through internal standards and goal systems, in Goal concepts in personality and social psychology Editor. Erlbaum: Hillsdale, NJ. pp. 19–85. (1989).
5. Bandura, A. Selective activation and disengagement of moral control. *J. Soc. Issues*. **46** (1), 27–46 (1990).
6. Detert, J. R., Trevino, L. K. & Sweitzer, V. L. Moral disengagement in ethical decision making: A study of antecedents and outcomes. *J. Appl. Psychol.* **93** (2), 374–391 (2008).
7. Newman, A. et al. Moral disengagement at work: A review and research agenda. *J. Bus. Ethics*. **167** (3), 535–570 (2020).
8. Schaefer, U. & Bouwmeester, O. Reconceptualizing moral disengagement as a process: transcending overly Liberal and overly Conservative practice in the field. *J. Bus. Ethics*. **172** (3), 525–543 (2020).
9. Moore, C. Moral disengagement. *Curr. Opin. Psychol.* **6**, 199–204 (2015).
10. Bandura, A. Moral disengagement in the Preparation of inhumanities. *Personality Social Psychol. Rev.* **3**, 193–209 (1999).
11. Welsh, D. T. et al. The slippery slope: how small ethical transgressions pave the way for larger future transgressions. *J. Appl. Psychol.* **100** (1), 114–127 (2015).
12. Tenbrunsel, A. E. & Messick, D. M. Ethical fading: the role of self-deception in unethical behavior. *Soc. Justice Res.* **17**, 223–236 (2004).
13. Cochran, W. & Tesser, A. *The What the Hell effect: Some effects of goal proximity and goal framing on performance*, in *Striving and Feeling: Interactions among Goals, Affect, and Self-regulation*, L.L. Martin and A. Tesser, Editors. Erlbaum: Mahwah, NJ. pp. 99–120. (1996).
14. Marlatt, G. A. *Relapse prevention: Theoretical rationale and overview of the model*, in *Relapse Prevention*, G.A. Marlatt and J.R. Gordon, Editors. Guilford: New York. pp. 3–70. (1985).
15. Loeber, R. et al. Initiation, escalation and desistance in juvenile offending and their correlates. *J. Criminal Law Criminol.* **82** (1), 36–82 (1991).
16. Mazar, N., Amir, O. & Ariely, D. The dishonesty of honest people: A theory of Self-Concept maintenance. *J. Mark. Res.* **45** (6), 633–644 (2008).
17. Effron, D. A. & Conway, P. When virtue leads to villainy: advances in research on moral self-licensing. *Curr. Opin. Psychol.* **6**, 32–35 (2015).
18. Jordan, J., Mullen, E. & Murnighan, J. K. Striving for the moral self: the effects of recalling past moral actions on future moral behavior. *Pers. Soc. Psychol. Bull.* **37**, 701–713 (2011).
19. Merritt, A. C., Effron, D. A. & Monin, B. Moral self-licensing: when being good frees Us to be bad. *Social Personality Psychol. Compass*. **4** (5), 344–357 (2010).
20. Zhong, C. B., Lijenquist, K. A. & Cain, D. M. *MORAL SELF-REGULATION licensing and Compensation*. Psychological perspectives on ethical behavior and decision making. ed. *D DeCremer* 75–89. (2009).
21. Kish-Gephart, J. et al. Situational moral disengagement: can the effects of self-interest be mitigated? *J. Bus. Ethics*. **125**, 267–285 (2014).

22. Gino, F. et al. Unable to resist temptation: how self-control depletion promotes unethical behavior. *Organ. Behav. Hum Decis. Process.* **115** (2), 191–203 (2011).
23. Lee, K. et al. Why victims of undermining at work become perpetrators of undermining: an integrative model. *J. Appl. Psychol.* **101**, 915–924 (2016).
24. Lin, S. H., Johnson, R. E. & J. Ma., and When ethical leader behaviors breaks bad: how ethical leader behavior can turn abusive via ego depletion and moral licensing. *J. Appl. Psychol.* **101**, 815–830 (2016).
25. Kouchaki, M. & Smith, I. H. The morning morality effect: the influence of time of day on unethical behavior. *Psychol. Sci.* **25** (1), 95–102 (2014).
26. Hystad, S. W., Mearns, K. J. & Eid, J. Moral disengagement as a mechanism between perceptions of organizational injustice and deviant work behaviors. *Saf. Sci.* **68**, 138–145 (2014).
27. Liu, Y. & Berry, C. M. Identity, moral, and equity perspectives on the relationship between experienced injustice and time theft. *J. Bus. Ethics*. **118**, 73–83 (2013).
28. Beaudoin, C. A., Cianci, A. M. & Tsakumis, G. T. *The impact of CFOs' incentives and earnings management ethics on their financial reporting decisions: The mediating role of moral disengagement* Journal of Business Ethics, 128: pp. 505–518. (2015).
29. Effron, D. A., Bryan, C. J. & Murnighan, J. K. Cheating at the end to avoid regret. *J. Personal. Soc. Psychol.* **109** (3), 395–414 (2015).
30. Garrett, N. et al. The brain adapts to dishonesty. *Nat. Neurosci.* **19** (12), 1727–1732 (2016).
31. Gino, F. & Bazerman, M. H. When misconduct goes unnoticed: the acceptability of gradual erosion in others' unethical behavior. *J. Exp. Soc. Psychol.* **45** (4), 708–719 (2009).
32. Moore, C. et al. Why employees do bad things: moral disengagement and unethical organizational behavior. *Pers. Psychol.* **65**, 1–48 (2012).
33. Mazar, N., Amir, O. & Ariely, D. More ways to Cheat - Expanding the scope of dishonesty. *J. Mark. Res.* **45** (6), 651–653 (2008).
34. Gino, F., Norton, M. I. & Ariely, D. The counterfeit self: the deceptive costs of faking it. *Psychol. Sci.* **21** (5), 712–720 (2010).
35. Mazar, N. & Zhong, C. B. Do green products make Us better people?? *Psychol. Sci.* **21** (4), 494–498 (2010).
36. Sharma, E. et al. Financial deprivation selectively shifts moral standards and compromises moral decisions. *Organ. Behav. Hum Decis. Process.* **123** (2), 90–100 (2014).
37. Barthelemy, S. & Boulinguez, P. Manual reaction time asymmetries in human subjects: the role of movement planning and attention. *Neurosci. Lett.* **315** (1–2), 41–44 (2001).
38. Mazar, N. & Hawkins, S. A. *Choice architecture in conflicts of interest: Defaults as physical and psychological barriers to (dis)honesty.* Journal of Experimental Social Psychology, 59: pp. 113–117. (2015).
39. Cheema, A. & Soman, D. The effect of partitions on controlling consumption. *J. Mark. Res.* **45** (6), 665–675 (2008).
40. Preacher, K. J. & Hayes, A. F. Asymptotic and resampling strategies for assessing and comparing indirect effects in multiple mediator models. *Behav. Res. Methods*. **40**, 879–891 (2008).
41. Mazar, N. & Ariely, D. Dishonesty in everyday life and its policy implications. *J. Public. Policy Mark.* **25** (1), 117–126 (2006).

Author contributions

CR and DA wrote the manuscript. Both authors reviewed the manuscript.

Declarations

Compliance with ethical standards

The authors declare no conflicts of interest. All participants provided informed consent and all procedures were approved by the Institutional Review Board. All methods were carried out in accordance with relevant guidelines and regulations.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-025-10097-9.

Correspondence and requests for materials should be addressed to C.R.