# Unethical amnesia responds more to instrumental than to hedonic motives

Fabio Galeotti, Charlotte Saucet, Marie Claire Villeval

# Unethical amnesia responds more to instrumental than to hedonic motives

Fabio Galeotti[a,1], Charlotte Saucet[b,1], and Marie Claire Villeval[a,c,1,2]

[a]Groupe d'Analyse et de Théorie Economique, UMR5824, Univ Lyon, CNRS, F-69130 Ecully, France; [b]CNRS, Sciences Po, 75337 Paris, France; and [c]Institute of Labor Economics, 53113 Bonn, Germany

Edited by Susan T. Fiske, Princeton University, Princeton, NJ, and approved August 30, 2020 (received for review June 3, 2020)

**Humans care about morality. Yet, they often engage in actions that contradict their moral self. Unethical amnesia is observed when people do not remember or remember less vividly these actions. This paper explores two reasons why individuals may experience unethical amnesia. Forgetting past unethical behavior may be motivated by purely hedonic or affective reasons, such as the willingness to maintain one's moral self-image, but also by instrumental or strategic motives, in anticipation of future misbehavior. In a large-scale incentivized online experiment ($n$ = 1,322) using a variant of a mind game, we find that hedonic considerations are not sufficient to motivate the forgetting of past cheating behavior. This is confirmed in a follow-up experiment ($n$ = 1,005) in which recalls are elicited the same day instead of 3 wk apart. However, when unethical amnesia can serve as a justification for a future action, such as deciding on whether to keep undeserved money, motivated forgetting is more likely. Thereby, we show that motivated forgetting occurs as a self-excuse to justify future immoral decisions.**

dishonesty | memory | motivated forgetting | unethical behavior | self-image

Examples of apparent amnesia abound in congressional testimonies or legal depositions. Claims of amnesia occur for murders (1), sexual harassment (2), domestic violence (3), and fraud (4). Often, amnesia serves as an attempt to avoid responsibility for past misdeeds, impede the investigation, or dodge embarrassing questions. However, memory impairments do not only concern criminals. They also affect ordinary people in their everyday lives. Although most people care about being seen as moral and perceive an intrinsic cost from cheating (5–9), they sometimes engage in actions that contradict their desire for a moral self-image. Each time people think about their past misdeeds, the moral costs of unethical behavior may reactivate and generate discomfort and distress, except if such memories can be obfuscated and distorted. If people are able to forget the details of their past transgressions, they can think of themselves as honest persons (10–12). The terms "unethical amnesia" (11), "motivated forgetting" (10) or, more broadly, "motivated memory" (13) have been coined to characterize the fact that individuals actively forget and shape the memory of their past unethical actions or unwanted experiences, so that their recalls become fuzzier as time passes. Motivated memory thus complements the wide range of strategies that people can adopt to preserve a positive self-image, such as avoiding information about the negative consequences of their behavior on others (14, 15), exploiting norm uncertainty (16), shifting the blame onto someone else (17, 18), balancing moral behavior over time (19–21), or using narratives to downplay the externalities of their actions or their pivotality (22).

However, we still know little about the reasons that motivate unethical amnesia. A first category of motives relates to purely hedonic moral cleansing, such as the regulation of affect, the justification of inappropriate behavior, and the preservation of image (10). Individuals who engaged in unethical behavior do not have to update their moral self-view if they are able to forget their past transgressions and therefore restore consistency between their memories and their moral self. A second possible motive is forward-looking, as unethical amnesia can serve as an instrument to justify future unethical actions. Indeed, memory manipulation can be used not only as postviolation justifications but also as "previolation justifications [that] lessen the *anticipated* threat to the moral self" (ref. 23, p. 1). For example, if individuals remember the eco-friendly actions they sometimes undertake but systematically forget the environmentally irresponsible ones, they will not only have a clean conscience from thinking of themselves as eco-friendly persons (and derive a positive utility from it) but they can also use these positive memories strategically to justify future irresponsible acts. These two mechanisms, hedonic and instrumental, are not mutually exclusive. While the first one is present-oriented and can develop in the absence of any future action, the second one is triggered by the anticipation of future decisions. This study investigates these two possible motives behind unethical amnesia.

Despite a vast literature on unethical behavior (for surveys see refs. 6 and 24–26) and a recent empirical psychological literature on unethical amnesia (10–12), memory manipulation as a self-management mechanism in which people can engage when facing a future opportunity to misbehave has remained almost unexplored. While theoretical economic models recognize both the affective and strategic values of motivated memory (27), empirical tests have focused only on the first one (11, 12, 28–32). In a setting where individuals are monetarily incentivized to recall their past actions accurately, we explore whether people manipulate their memory both to sustain their desire for moral

## Significance

**Using large-scale incentivized online experiments, we tested two possible origins of individuals' forgetting about their past cheating behavior in a mind game. We found that purely hedonic considerations, such as the maintenance of a positive self-image, are not sufficient to motivate unethical amnesia, but the addition of an instrumental value to forgetting triggers such amnesia. Individuals forget their past lies more when amnesia can serve as an excuse not to engage in future morally responsible behavior. These findings shed light on the interplay between dishonesty and memory and suggest further investigations of the cost function of unethical amnesia. A policy implication is that improving ethics requires making unethical amnesia more difficult for individuals.**

Author contributions: F.G., C.S., and M.C.V. designed research, performed research, analyzed data, and wrote the paper.

The authors declare no competing interest.

This article is a PNAS Direct Submission.



[1]F.G., C.S., and M.C.V. contributed equally to this work.

[2]To whom correspondence may be addressed. Email: villeval@gate.cnrs.fr.

This article contains supporting information online at https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2011291117/-/DCSupplemental.

First published September 28, 2020.

self-image, and/or as an excuse not to engage in subsequent morally responsible behavior (giving back some undeserved money). Our study presents an experimental test of the impact of anticipated decisions on memory manipulation. It complements previous work (11, 12, 33) that, in contrast to us, investigated memory biases as a consequence of past unethical behavior, relied on attitudinal measures of memory rather than on behavior, and did not incentivize participants for providing truthful recalls.

## Testing Hedonic and Instrumental Motives behind Unethical Amnesia

To study 1) whether people tend to forget their past dishonest decisions and 2) whether this results more from the maintenance of moral self-image or from using memory as an instrument to justify future decisions, we ran a large-scale two-part online experiment ($n$ = 1,322) (see details in *SI Appendix*, section A1 and Table S1).

In the first part of the experiment, we designed a variant of a mind game (8, 34–37) inspired by cheating games (5, 38). Participants played 20 rounds of a "wheel game" where they could misreport the outcome of a random draw to increase their payoff at no risk of detection. The program displayed a wheel with six empty squares, and participants were asked to choose one square in their head. Then, the program randomly displayed a number between 1 and 6 in each square of the wheel. Each number appeared only once. Participants were asked to report the number displayed in the square they had previously chosen. Participants' earnings for this part were determined by the number they reported in one randomly selected round. The higher the reported number, the higher the earnings.

In the second part of the experiment, conducted 3 wk later, participants were first informed that, depending on a random draw, they might be given the possibility to reduce their payoff. If given this possibility, they would have to decide at the end of the experiment whether or not to reduce their payoff by a fixed amount, and they were encouraged to do so if they had misreported several numbers to their advantage in the wheel game. All participants were told about the two conditions, so that the salience of the moral dimension was kept constant across conditions. Then, participants were informed about their own condition, and they learned that in this part they would be asked to recall the distribution of the 20 numbers they reported in the first part of the experiment. Precisely, they had to recall how many times they reported the numbers 1, 2, 3, 4, 5, and 6.

We varied between subjects whether only hedonic or both hedonic and instrumental reasons could motivate memory. The Hedonic treatment corresponds to the condition in which participants ($n$ = 488) learned before performing the memory task that they were not given the option to reduce their payoff at the end of the experiment. Thus, biased memory errors, if any, result from purely affective reasons. The Instrumental treatment corresponds to the condition in which participants ($n$ = 508) learned before performing the memory task that they were given the possibility to reduce their payoff at the end, and encouraged to do so if they overreported their outcomes. Biased memory errors in this treatment may be motivated not only by the maintenance of moral image but also by strategic reasons, that is, not giving back part of the undeserved money by persuading oneself that one is not concerned.

Since accurate recalls were incentivized, any memory errors from honest participants can be attributed to random rather than motivated forgetting, and there should be as many errors below as above the actual numbers. However, it might be easier to recall a uniform or close-to-uniform distribution rather than a biased distribution. Thus, dishonest individuals could display more memory errors not because of motivated memory but because they had to recall a different distribution relative to honest individuals. We thus ran two control treatments: Hedonic-Control ($n$ = 163) and Instrumental-Control ($n$ = 163). The only difference with the previous treatments is that participants could not cheat in the first part. They had to choose one empty square in each wheel by clicking on it instead of choosing it in their head before the numbers were displayed on the wheel, and they could not report a different number than the one displayed in the chosen square. To allow comparisons, the numbers displayed on the chosen squares (not wheels) reflected the aggregate distribution of the numbers reported in the respective Hedonic and Instrumental treatments. In these control treatments, participants had no reason to manipulate their memory. The comparison of the main treatments with the control treatments allows us to separate motivated from nonmotivated forgetting.

**Measuring Dishonesty.** To identify the individuals who misreported numbers to their advantage in the wheel task of part 1 in the two main treatments, each individual's distribution of reported numbers is compared with a uniform distribution (i.e., the expected distribution of a participant who truthfully reported all numbers). A participant is classified as dishonest if the average report is greater than 3.5 (i.e., the mean of a uniform distribution) and the reported distribution differs significantly from the uniform at the 10% level ($\chi^2$ test). Otherwise, the participant is classified as honest. We chose a 10% threshold to classify participants as honest or dishonest instead of 5% because a stricter threshold would lead us to classify as honest individual participants whose distribution of reported numbers is farther from the uniform distribution. There is a trade-off between the threshold used to define the two categories of participants and the size of the difference between the two distributions. We favored the 10% threshold to generate a larger difference between the two distributions and provide a better test of the difference in recalls between a more uniform and a less uniform distribution. Also, setting a threshold at 10% is equivalent to conducting a one-sided test, and this is justified since we only consider left-skewed distributions (by imposing that the average report is greater than 3.5) in the definition of a dishonest subject. Note that our results are robust to classifications using different thresholds (see *SI Appendix*, section A3).

The same method serves to separate participants in the control treatments. They are classified as having to recall a nonuniform distribution if the average of the numbers displayed in their selected squares is greater than 3.5 and their distribution differs significantly from the uniform distribution at the 10% level ($\chi^2$ test). Otherwise, they are classified as having to recall an almost-uniform distribution ("almost" because they did not see exactly the same frequency for each number in the selected squares). Note that we also impose a condition on the mean number to be able to compare the memory of participants classified as dishonest in the main treatments and that of participants having to recall a nonuniform distribution in the control treatments. Overall, 27.25% (26.77%) of the participants from the Hedonic (Instrumental) treatment are classified as dishonest individuals, and 30.06% (30.06%) of the participants from the Hedonic-Control (Instrumental-Control) treatment had to recall a nonuniform distribution. Descriptive statistics on participants' cheating behavior are provided in *SI Appendix*, section A4.

Some dishonest participants cheated to the full extent by always reporting the highest number. We classify these participants as "full cheaters." The remaining dishonest participants are classified as "partial cheaters." Out of 996 participants allocated to the Hedonic and Instrumental treatments, 55 cheated to

the full extent. We analyze separately the behavior of partial and full cheaters for several reasons. First, in the control treatments, none of the subjects had to recall twenty 6s. Hence, excluding the full cheaters allows a clean comparison between the control and main treatments. Second, partial and full cheaters are likely to differ in their very nature: Full cheaters may care less about their image and may not feel the need to manipulate their memory. Finally, the effort that partial and full cheaters have to provide to recall their behavior may not be the same: it may be easier to recall twenty 6s than different numbers. Below, we first report the results excluding full cheaters and then briefly comment on the differences between full and partial cheaters. A more detailed analysis of the behavior of partial and full cheaters, with additional robustness checks, is provided in *SI Appendix*, sections A5 and A6.

**Measuring Memory Errors.** Participants' memory errors are defined as the difference between the average reported outcome in the first part of the experiment and the average recalled outcome in the second part. While a negative memory error is exclusively nonmotivated, a positive memory error may capture both motivated and nonmotivated forgetting. A positive average memory error thus captures motivated memory.

## Results

**Memory Errors When Actions Are Not Unethical.** In the Control treatments, memory errors result from cognitive limitations. Manipulating the conditions that would trigger different motives for unethical amnesia has naturally no impact on such errors. This is attested by Fig. 1, which displays the average memory error in the Hedonic-Control and Instrumental-Control treatments, both for participants who had to recall an almost-uniform distribution (Fig. 1, *Left*) and those who had to recall a nonuniform distribution (Fig. 1, *Right*). It shows that the average memory error of individuals who had to recall an almost-uniform distribution is 0.11 in the Instrumental-Control treatment and 0.08 in the Hedonic-Control treatment, which is statistically indistinguishable ($P = 0.764$, Wald test). The average memory error of individuals who had to recall a nonuniform distribution is 0.28 in the Instrumental-Control treatment and 0.29 in the Hedonic-Control treatment, which is again not significantly different ($P = 0.904$). There is no difference either between the Hedonic-Control and Instrumental-Control treatments when pooling all participants together ($Mean_{ME} = 0.15$ and 0.16 in the Hedonic-Control and Instrumental-Control treatments, respectively; $P = 0.854$). Therefore, we pool the control treatments for the remainder of the analysis and refer to them as "Pooled-Control." Note that Fig. 1 also shows that it is more difficult to recall a nonuniform distribution than an almost-uniform one ($P = 0.008$) even for participants who had no chance to misreport their outcomes.

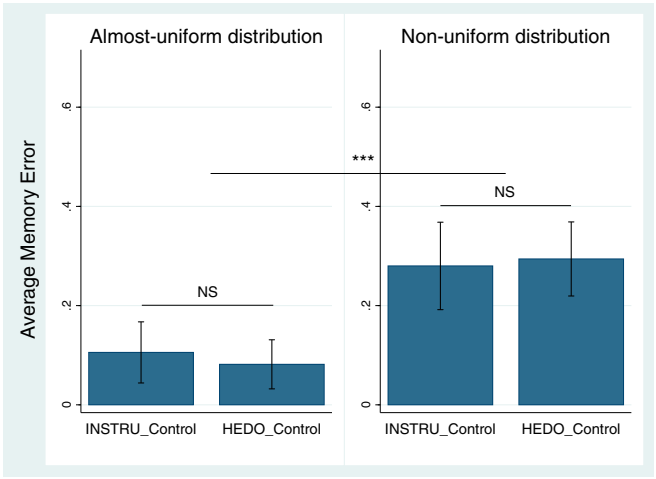


**Fig. 1.** Average memory error in the control treatments. INSTRU is for Instrumental and HEDO for Hedonic. In the controls, participants could not cheat but faced numbers reflecting a nonuniform or an almost-uniform aggregate distribution of the numbers reported in the respective Hedonic and Instrumental treatments. Participants do not forget significantly more in the Instrumental-Control than in the Hedonic-Control treatment. Having to recall a nonuniform distribution (*Right*) leads to more memory errors than having to recall an almost-uniform distribution (*Left*). Significance levels are computed from Model 1 (almost-uniform distributions), Model 4 (nonuniform distributions), and Model 7 (almost-uniform vs. nonuniform distributions) of *SI Appendix*, Table S2 (see also *SI Appendix*, section A7). Error bars indicate standard errors of the mean. See main text for statistics. ***$P < 0.01$, NS = not significant.

Turning to the main treatments, there is no evidence of motivated amnesia when participants reported honestly. Fig. 2, *Left* shows no significant difference in average memory errors between participants classified as honest in the Hedonic and Instrumental treatments ($Mean_{ME}$ = 0.02 and 0.04, respectively; $P = 0.593$, Wald test). There is no difference either when the participants classified as honest are compared to participants who could not cheat but had to recall an almost-uniform distribution in the Pooled-Control treatment ($Mean_{ME}$ = 0.09; $P = 0.105$ and $P = 0.245$, respectively) (see *SI Appendix*, section A5 and Table S4).

**Instrumental Amnesia Is the Main Source of Motivated Memory.** Comparing the average memory error of participants who had to recall a nonuniform distribution in the Hedonic treatment and the Pooled-Control treatment indicates whether dishonest individuals exhibit motivated forgetting when it allows them to maintain their moral image. Comparing the Instrumental treatment and the Pooled-Control and the Hedonic treatments, respectively, indicates whether unethical amnesia is more important when forgetting can be used as a self-justification to not give undeserved money back.

Fig. 2, *Right* shows that dishonest participants exhibit higher memory errors in the Instrumental treatment ($Mean_{ME} = 0.54$) than in the Hedonic treatment ($Mean_{ME} = 0.29$; $P = 0.006$, Wald test) and than individuals who had to recall a nonuniform distribution in the Pooled-Control treatment ($Mean_{ME} = 0.29$; $P = 0.006$). By contrast, Fig. 1, *Right* shows that there was no difference in errors between Instrumental-Control and Hedonic-Control for individuals who had to recall a nonuniform distribution but could not cheat. In the Hedonic treatment, the average memory error of dishonest individuals is indistinguishable from the average error of the participants who had to recall a nonuniform distribution in the Pooled-Control treatment ($P = 0.963$). These results are supported by the regression analyses reported in *SI Appendix*, Table S3 (also see *SI Appendix*, section A5), including various control variables. All models of *SI Appendix*, Table S3 show that dishonest individuals in the Instrumental treatment exhibit significantly higher memory errors than dishonest individuals in the Hedonic treatment, whereas the latter do not exhibit higher memory errors than individuals who had to recall a nonuniform distribution in the Pooled-Control treatment. Thus, dishonest individuals exhibit higher memory errors when an instrumental motive for forgetting is present, that is, when it serves as a self-justification to keep undeserved money, while purely hedonic motives are not sufficient in our study to generate unethical amnesia.

**Full Cheaters Are Less in Need of Instrumental Amnesia.** There is evidence of instrumental forgetting for partial cheaters, but what about full cheaters (those who reported a "6" twenty times in the wheel task)? In the Instrumental treatment, the average

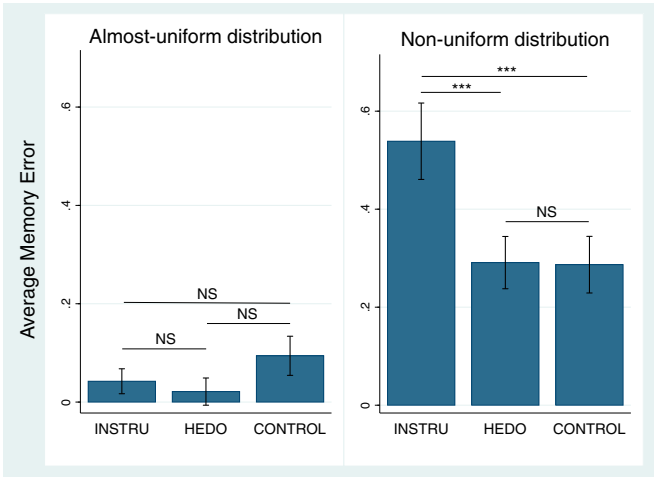


**Fig. 2.** Average memory error of the honest and dishonest participants in the main treatments. INSTRU is for Instrumental, HEDO for Hedonic, and CONTROL for the Pooled-Control treatment. The average memory error of honest participants in the main treatments does not differ from the average error of participants who had to recall an almost-uniform distribution in the Pooled-Control treatment (*Left*). The average memory error of dishonest participants in the Hedonic treatment does not differ from that of participants who had to recall a nonuniform distribution in the Pooled-Control treatment (*Right*). However, the average error of dishonest participants in the Instrumental treatment differs from the average error of those who had to recall a nonuniform distribution in the Pooled-Control treatment. Significance levels are computed from Model 1 in *SI Appendix*, Table S3 (for dishonest subjects) and *SI Appendix*, Table S4 (for honest subjects) (see also *SI Appendix*, section A5). Full cheaters are not included. Error bars indicate standard errors of the mean. ***$P < 0.01$, NS = not significant.

memory error of full cheaters ($n = 29$, $Mean_{ME} = 0.26$) is significantly lower than that of partial cheaters ($n = 88$, $Mean_{ME} = 0.54$) ($P < 0.001$, Mann–Whitney $U$ test). This does not result from differences in cognitive memory abilities between these two groups, as no difference in performance is observed between partial and full cheaters in a word memory task based on ref. 39 that we administered at the end of part 1. On average and on aggregate, partial cheaters recalled correctly 25.58 (25.86 in the Instrumental treatment) words out of 35 and full cheaters 24.63 (24.24) ($P = 0.426$ and 0.510, Mann–Whitney $U$ tests). Also, the difference does not seem to be driven by the fact that recalling is cognitively easier for full cheaters than for partial cheaters. Indeed, in the Hedonic treatment, there is no significant difference in the mean memory error between full cheaters and partial cheaters ($Mean_{ME} = 0.39$ and 0.29, respectively, $P = 0.448$). Moreover, for the whole population, the correlation between the average memory error and the SD of the numbers actually reported in the wheel task is significant neither at the aggregate level ($\rho = -0.015$, $P = 0.792$, pairwise Pearson's correlation), nor at the treatment level ($\rho = -0.020$, 0.040, and 0.050; $P = 0.858$, 0.674, and 0.595 in the Pooled-Control, Hedonic, and Instrumental treatments, respectively). The regression analysis reported in *SI Appendix*, Table S3 confirms that a lower SD in the distribution of the numbers reported does not significantly decrease memory errors. Hence, the higher memory errors exhibited by partial cheaters does not seem to be explained by the higher variation in the numbers they had to recall. A more likely explanation for the difference in memory errors between full and partial cheaters in the Instrumental treatment is that these two types of individuals differ in terms of intrinsic motivation. Full cheaters, who deliberately chose to maximize their payoff in the wheel task, might care less about their image and, thus, may not need additional arguments to not return part of the undeserved money. By contrast, partial cheaters most likely faced a trade-off in part 1 between overreporting the numbers to increase their earnings and maintaining a positive self-image. For these participants, forgetting their past cheating behavior could have been the only viable self-management strategy to self-justify not returning the undeserved money.

**Hedonic Motives Do Not Affect the Quality of Short-Term Memory Either.** Finding no difference in average memory errors between the Hedonic and the Pooled-Control treatments indicates that affective motives alone are not sufficient to trigger motivated memory. However, participants in the control treatments reported numbers that they did not get to choose. Thus, memory errors in these treatments may also capture inattention, which may have produced more volatile souvenirs than when people were actively dishonest. This would be consistent with Saucet and Villeval (30), who found more memory errors in dictator games when the amount to recall was randomly assigned by the program than when it was chosen by the dictator. Therefore, we conducted a follow-up experiment ($n = 1{,}005$) (see details in *SI Appendix*, section A1 and Table S9). The Short-Term-Memory treatments replicated the same treatments as in the original experiment, except that the wheel game and the recall task were performed in the same session. Thus, even if participants were possibly still less attentive in the control than in the main treatments, they had a higher chance to recall their reported number distribution. In this setting, the comparison between the new Hedonic and control treatments (all indexed by ST for Short-Term-Memory) provides additional insight on the existence of an affective motive behind motivated memory.

Not surprisingly, in each Short-Term-Memory treatment the average memory error is lower than in our original experiment because there is less time to forget. Like in the original experiment, there is no memory bias when participants reported honestly (Fig. 3, *Left* and *SI Appendix*, Table S11). None of the treatment pairwise comparisons are significant ($P > 0.10$). By contrast, when considering dishonest participants (Fig. 3, *Right*),

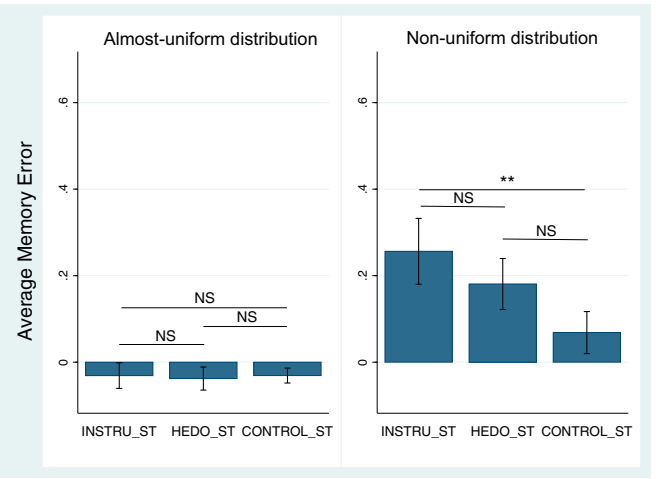


**Fig. 3.** Average memory error of the honest and dishonest participants in the Short-Term-Memory (ST) treatments. The average memory error of honest participants in the ST treatments does not differ from that of participants who had to recall an almost-uniform distribution in the Pooled_Control_ST treatment (*Left*). Only the average memory error of dishonest participants in the Instrumental_ST treatment differs from the average error of participants who had to recall a nonuniform distribution in the Pooled-Control_ST treatment (*Right*). Significance levels are computed from Model 1 of *SI Appendix*, Table S10 (dishonest subjects) and *SI Appendix*, Table S11 (honest subjects). Full cheaters are not included. Error bars indicate standard errors of the mean. **$P < 0.05$, NS = not significant.

a ranking appears clearly: The average memory error is the highest in the Instrumental_ST treatment ($Mean_{ME} = 0.26$), followed by the Hedonic_ST treatment ($Mean_{ME} = 0.18$), and by the Pooled-Control_ST treatment ($Mean_{ME} = 0.07$). The regression analysis reported in *SI Appendix*, Table S10 and summarized in *SI Appendix*, section A9 indicates that the mean memory errors in the Instrumental_ST treatment differ from the control ($P = 0.029$ in Model 1, $P = 0.054$ in Model 4), showing some evidence of instrumental forgetting also in the short term. However, the Hedonic_ST treatment is not significantly different from the control ($P > 0.10$ in all models), and most of the time the sign of its coefficient is even negative. Thus, hedonic motives for memory distortion are not sufficiently strong to generate a statistically significant difference in the quality of short-term recalls compared to that of individuals who had no opportunity to cheat. It may be more difficult to manipulate one's memory when the retrieval of a decision is closer in time to the realization of this decision, even when one has a strategic reason to do so (not returning money).

## Discussion

Besides providing clean evidence of motivated forgetting in the context of dishonest decision making, our study investigated two possible motives behind unethical amnesia. We showed that hedonic motives are not sufficient to motivate forgetting of past cheating behavior. It is only when unethical amnesia has an instrumental value—that is, when forgetting can serve as a self-justification for a future action—that individuals who misbehaved in the past motivated their memory. This important result indicates that people experience motivated memory not necessarily to restore their moral self-image after a misconduct but more as a self-excuse to justify future immoral decisions.

Indeed, individuals who behaved dishonestly in the Hedonic treatment did not exhibit significantly higher memory errors than those who had to recall a similar nonuniform distribution but did not engage in any moral decision before. This absence of hedonic motive behind unethical amnesia when no future decision has to be made does not necessarily contradict previous evidence (11, 12, 33). First, even our Short-Term-Memory treatments provides a conservative test because of the higher passiveness of participants in the control treatments, which may have impacted the quality of memory negatively. Second, most of the previous experiments investigating memories of cheating did not incentivize truthful recalls, whereas our participants received monetary incentives for accurate recalls. Providing sufficiently high monetary incentives for recalls may induce people to recover image-threatening memory traces that they had initially suppressed (32). Monetary utility might somewhat compensate for a loss in ego-utility.

While purely hedonic reasons were insufficient to generate unethical amnesia, the addition of an instrumental motive triggered self-serving forgetting. Dishonest individuals recalled their past behavior with less accuracy when being informed that they would have a future decision to make that engaged again their morality. By contrast, when individuals did not have to engage in prior ethical decision making, the addition of a future moral decision did not affect the quality of recalls. A possible reason for these results is that people underestimated their past unethical behavior because they needed self-justification for keeping undeserved money.

While previous empirical studies investigated memory manipulation as a present-oriented strategy to cope with past self-image threatening decisions, we show that it is crucial to consider memory manipulation also as a future-oriented self-management tool. Our results also complement prior research demonstrating that people are more likely to act unethically repeatedly over time because they experience unethical amnesia (11). Our study shows that the reverse causality is also true: People experience unethical amnesia when this can serve as an excuse to justify future unethical behavior.

More empirical evidence is needed to establish whether motivated forgetting is more or less deliberate. While some theoretical models consider that individuals are naive and memory biases are the result of a heuristic process (40–43), other models grant large metacognitive control to individuals, so that forgetting is active and deliberate (27, 44). Confronting these models empirically might require to combine experimental methods with neuroimaging (45) and neurophysiology techniques (46). If individuals deliberately inhibit control over encoding or retrieval, which would involve a more controlled computation of costs and benefits, such memory manipulation might activate different brain areas and neural networks, compared to genuine forgetting. These networks might also differ depending on whether unethical amnesia is present- or future-oriented. Conducting such studies would nicely contribute to the debate on metacognition and individuals' degree of sophistication. It would be also important to investigate the characteristics of the function that relates the moral cost of unethical behavior with the intensity of unethical amnesia. In our experiment partial cheaters exhibited more strategic unethical amnesia than full cheaters. The moral costs of lying prevent most individuals from fully exploiting their cheating opportunities (5–7, 9) but these costs may be reduced if individuals can anticipate that they are able to manipulate their memories. However, while there is a debate on the moral cost function of cheating (8, 37), the cost function of unethical amnesia is totally unknown. Exploring to which extent making unethical amnesia more costly would translate into more ethical behavior would be particularly helpful to infer policy recommendations.

## Materials and Methods

Details of the experimental design and procedures and tables can be found in *SI Appendix*, section A1. All of the participants gave their informed consent at the beginning of the experiment. The Review Board of the Groupe d'Analyse et de Théorie Economique at the University of Lyon reviewed and approved the procedures (2018-0015).

Individuals located in the United States were recruited through Amazon Mechanical Turk to participate in our two-part online study and were informed that they could complete the second part of the study 3 wk after the first part; 1,322 individuals (659 males; mean age = 39.58 y, SD = 11.22 y) completed the two parts (1,550 completed the first part); 488 individuals participated in the Hedonic treatment (235 males; mean age = 40.01 y, SD = 11.40 y), 508 in the Instrumental treatment (275 males; mean age = 40.23 y, SD = 11.22 y), 163 in the Hedonic-Control treatment (74 males; mean age = 38.21 y, SD = 10.93 y), and 163 in the Instrumental-Control treatment (75 males; mean age = 37.62 y, SD = 10.68 y). The statistical power analysis is described in *SI Appendix*, section A2.

No feedback on performance or earnings was provided until all of the two parts were completed. At the end of the first part, participants received a fixed payoff of \$1.5 for their participation. At the end of the second part, they received a fixed payoff of \$1.5 for completing the second part, plus the joint earnings made in the first and second parts to make sure that most participants would return to complete the second part. They earned on average \$4.31 (SD = 0.48).

Participants in the follow-up Short-Term-Memory experiment were also located in the United States and were recruited through Amazon Mechanical Turk; 1,005 individuals (597 males; mean age = 36.87 y, SD = 11.24 y) participated in total, 254 in the Hedonic_ST treatment (156 males; mean age = 36.13 y, SD = 10.31 y), 251 in the Instrumental_ST treatment (147 males; mean age = 37.77 y, SD = 11.44 y), 252 in the Hedonic-Control_ST treatment (152 males; mean age = 36.72 y, SD = 12.08 y), and 248 in the Instrumental-Control_ST treatment (142 males; mean age = 36.88 y, SD = 11.05 y). They earned on average \$4.35 (SD = 0.48).

**Data Availability.** Data have been deposited in Open Science Framework (doi: 10.17605/OSF.IO/QMC8J).

**ACKNOWLEDGMENTS.** We are grateful to P. Seabright, S. Suetens, B. Tarroux, and F. Zimmermann for their very helpful comments and seminar participants at the 2019 European Economic Science Association meeting

in Dijon for their feedback. We also thank MTurk Data for assisting with data collection. This research has been funded by the Laboratory of Excellence CORTEX (construction, cognitive function, rehabilitation and repair of the cortex) (ANR-11-LABX-0042) of Université de Lyon within the program "Investissements d'Avenir" (ANR-11-IDEX-007) operated by the French National Research Agency and by the DECISION project (A study on Dishonesty: from Emotions and Cognition to Institutions and Organizations) of the French National Research Agency (ANR-19-CE26-0019). The research has benefited from the support of Initiative of Excellence Lyon (IDEXLYON) from Université de Lyon (project Institutional Design and Economic Preferences: Theory and Experiments - INDEPTH) within the Program Investissements d'Avenir (ANR-16-IDEX-0005). C.S. gratefully acknowledges the European Research Council, grant 850996, MOREV (Motivated Reading of Evidence) project.

1. P. J. Taylor, M. D. Kopelman, Amnesia for criminal offences. *Psychol. Med.* **14**, 581–588 (1984).
2. D. Bourget, J. M. Bradford, Sex offenders who claim amnesia for their alleged offense. *Bull. Am. Acad. Psychiatry Law* **23**, 299–307 (1995).
3. G. Swihart, J. Yuille, S. Porter, The role of state-dependent memory in "red-outs". *Int. J. Law Psychiatry* **22**, 199–212 (1999).
4. M. D. Kopelman, R. Green, E. Green, P. Lewis, N. Stanhope, The case of the amnesic intelligence officer. *Psychol. Med.* **24**, 1037–1045 (1994).
5. U. Fischbacher, F. Föllmi-Heusi, Lies in disguise—An experimental study on cheating. *J. Eur. Econ. Assoc.* **11**, 525–547 (2013).
6. J. Abeler, D. Nosenzo, C. Raymond, Preferences for truth-telling. *Econometrica* **87**, 1115–1153 (2019).
7. M. Dufwenberg, M. A. Dufwenberg, Lies in disguise–A theoretical analysis of cheating. *J. Econ. Theory* **175**, 248–264 (2018).
8. U. Gneezy, A. Kajackaite, J. Sobel, Lying aversion and the size of the lie. *Am. Econ. Rev.* **108**, 419–453 (2018).
9. K. Khalmetski, D. Sliwka, Disguising lies–Image concerns and partial lying in cheating games. *Am. Econ. J. Microecon.* **11**, 79–110 (2019).
10. M. Kouchaki, F. Gino, Dirty deeds unwanted: The use of biased memory processes in the context of ethics. *Curr. Opin. Psycol.* **6**, 82–86 (2015).
11. M. Kouchaki, F. Gino, Memories of unethical actions become obfuscated over time. *Proc. Natl. Acad. Sci. U.S.A.* **113**, 6166–6171 (2016).
12. M. L. Stanley, P. Henne, V. Iyengar, W. Sinnott-Armstrong, F. De Brigard, I'm not the person I used to be: The self and autobiographical memories of immoral actions. *J. Exp. Psychol. Gen.* **146**, 884–895 (2017).
13. J. A. Singer, P. Salovey, "Motivated memory: Self-defining memories, goals, and affect regulation" in *Striving and Feeling: Interactions Among Goals, Affect, and Self-Regulation*, L. L. Martin, A. Tesser, Eds. (Psychology Press, 1996), pp. 229–250.
14. L. Feiler, Testing models of information avoidance with binary choice dictator games. *J. Econ. Psychol.* **45**, 253–267 (2014).
15. Z. Grossman, J. J. Van Der Weele, Self-image and willful ignorance in social decisions. *J. Eur. Econ. Assoc.* **15**, 173–217 (2017).
16. C. Bicchieri, E. Dimant, S. Sonderegger, It's not a lie if you believe it: Lying and belief distortion under norm-uncertainty. *SSRN* (2019). https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3326146. Accessed 15 September 2020.
17. B. Bartling, U. Fischbacher, Shifting the blame: On delegation and responsibility. *Rev. Econ. Stud.* **79**, 67–87 (2011).
18. R. Oexl, Z. J. Grossman, Shifting the blame to a powerless intermediary. *Exp. Econ.* **16**, 306–312 (2013).
19. M. Ploner, T. Regner, Self-image and moral balancing: An experimental analysis. *J. Econ. Behav. Organ.* **93**, 374–383 (2013).
20. U. Gneezy, A. Imas, K. Madarász, Conscience accounting: Emotion dynamics and social behavior. *Manage. Sci.* **60**, 2645–2658 (2014).
21. D. Cojoc, A. Stoian, Dishonesty and charitable behavior. *Exp. Econ.* **17**, 717–732 (2014).
22. R. Bénabou, A. Falk, J. Tirole, "Narratives, imperatives and moral reasoning" (CEPR Discussion Paper 13056, Center for Economic and Policy Research, Washington, DC, 2018).
23. S. Shalvi, F. Gino, R. Barkan, S. Ayal, Self-serving justifications: Doing wrong and feeling moral. *Curr. Dir. Psychol. Sci.* **24**, 125–130 (2015).
24. S. M. Rosenbaum, S. Billinger, N. Stieglitz, Let's be honest: A review of experimental evidence of honesty and truth-telling. *J. Econ. Psychol.* **45**, 181–196 (2014).
25. B. Irlenbusch, M. C. Villeval, Behavioral ethics: How psychology influenced economics and how economics might inform psychology?. *Curr. Opin. Psycol.* **6**, 87–92 (2015).
26. C. Jacobsen, T. R. Fosgaard, D. Pascual-Ezama, Why do we lie? A practical guide to the dishonesty literature. *J. Econ. Surv.* **32**, 357–387 (2018).
27. R. Bénabou, J. Tirole, Self-confidence and personal motivation. *Q. J. Econ.* **117**, 871–915 (2002).
28. K. K. Li, "What determines overconfidence and memory recall bias? The role of feedback, awareness and social comparison" (Mimeo, City University of Hong Kong, 2017).
29. S. H. Chew, W. Huang, X. Zhao, Motivated false memory. *SSRN* (2020). https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2127795. Accessed 15 September 2020.
30. C. Saucet, M. C. Villeval, Motivated memory in dictator games. *Game. Econ. Behav.* **117**, 250–275 (2019).
31. R. W. Carlson, M. A. Maréchal, B. Oud, E. Fehr, M. J. Crockett, Motivated misremembering of selfish decisions. *Nat. Commun.* **11**, 2100 (2020).
32. F. Zimmermann, The dynamics of motivated beliefs. *Am. Econ. Rev.* **110**, 337–361 (2020).
33. L. L. Shu, F. Gino, M. H. Bazerman, Dishonest deed, clear conscience: When cheating leads to moral disengagement and motivated forgetting. *Pers. Soc. Psychol. Bull.* **37**, 330–349 (2011).
34. T. Jiang, Cheating in mind games: The subtlety of rules matters. *J. Econ. Behav. Organ.* **93**, 328–336 (2013).
35. S. Shalvi, C. K. De Dreu, Oxytocin promotes group-serving dishonesty. *Proc. Natl. Acad. Sci. U.S.A.* **111**, 5503–5507 (2014).
36. J. Potters, J. Stoop, Do cheaters in the lab also cheat in the field? *Eur. Econ. Rev.* **87**, 26–33 (2016).
37. A. Kajackaite, U. Gneezy, Incentives and cheating. *Game. Econ. Behav.* **102**, 433–444 (2017).
38. S. Shalvi, J. Dana, M. J. Handgraaf, C. K. De Dreu, Justified. ethicality: Observing desired counterfactuals modifies ethical perceptions and behavior. *Organ. Behav. Hum. Decis. Process.* **115**, 181–190 (2011).
39. H. L. Roediger, K. B. McDermott, Creating false memories: Remembering words not presented in lists. *J. Exp. Psychol. Learn. Mem. Cognit.* **21**, 803–814 (1995).
40. D. Kahneman, A. Tversky, Subjective probability: A judgment of representativeness. *Cognit. Psychol.* **3**, 430–454 (1972).
41. S. Mullainathan, A memory-based model of bounded rationality. *Q. J. Econ.* **117**, 735–774 (2002).
42. P. Bordalo, N. Gennaioli, A. Shleifer, "Memory, attention, and choice" (NBER Working Paper No. 23256, National Bureau of Economic Research, Cambridge, MA, 2017).
43. A. L. Bodoh-Creed, Mood, memory, and the evaluation of asset prices. *Rev. Finance* **24**, 227–262 (2020).
44. D. Gottlieb, Imperfect memory and choice under risk. *Game. Econ. Behav.* **85**, 127–158 (2014).
45. M. Anderson, S. Hanslmayr, Neural mechanisms of motivated forgetting. *Trends Cognit. Sci.* **18**, 279–292 (2014).
46. X. Hu, Z. Bergstrom, G. Bodenhausen, J. Rosenfeld, Suppressing unwanted autobiographical memories reduces their automatic influences: Evidence from electrophysiology and an implicit autobiographical memory test. *Psychol. Sci.* **26**, 1098–1106 (2015).