“I would not have done that”: Outcome knowledge distorts memory for decisions made minutes ago

Keishiro Sawa[1], Carla Zoe Cremer[1], Tobias Gerstenberg[2], Sanjay Manohar[1] & Matan Mazor[1]
[1]University of Oxford
[2]Stanford University

Abstract

Remembering one’s past decisions is critical for learning from delayed feedback and maintaining an accurate self-model. Across two gamified experimental paradigms and three experiments (total N = 500), we find that a selective forgetting of unsuccessful guesses operates within minutes, that it evades introspection, and that it affects confidence ratings in memory judgments. Analysis of memory errors reveals that, in reconstructing their past decisions, people integrate what they remember deciding, what they believe they would have decided, and what they would decide now if asked to make the same decision again.

Keywords: episodic memory; metacognition; hindsight bias; self-model

Introduction

Episodic memory is a record not only of external events, but also of one’s own actions and decisions. For example, a son’s episodic memory of a phone call he had with his mother includes information not only about what his mother said in the call, but also about what he himself told her, and how he responded to what she told him. The transcript of the conversation will normally not be encoded in memory in verbatim form: episodic memories can be efficiently represented by relying on simplified models or schemas (Bartlett, 1995; Hemmer & Steyvers, 2009; Levin, 2024; Spens & Burgess, 2024). For example, the son may accurately fill in missing details from his mental record of the conversation by relying on his mental model of his mother and the kinds of things she might say. Crucially, the same is true for his mental record of what he himself said in the conversation: he may be able to reconstruct his words by relying on his self-model, which specifies the kind of things he is likely to do and say under different circumstances (Mazor, Firestone, & Phillips, 2026; Mazor, Seghezzi, & Manohar, 2026).

Despite its crucial importance for understanding episodic memory, little research has looked into people’s memory of their own actions and decisions, perhaps because it is challenging to experimentally study. Participants’ success in remembering a list of arbitrary words, or details from a video clip, necessarily means that they successfully encoded and retrieved this information. In contrast, successfully reproducing the same behaviour again can mean that the information was retained in memory, but it can also mean that it was reconstructed based on a good self-model (“I don’t remember what I did, but this is what I *would* do”), or generated anew from the same decision policy (“I don’t remember what I did, but this is what I *should* have done”). Crucially, these same aspects that make memory for self actions difficult to study also make it an important target for scientific investigation, as the validity of these additional sources of evidence will have important implications for people’s ability to remember their actions. This ability will, in turn, be critical for learning from delayed feedback (Biderman et al., 2020; Gershman & Daw, 2017; Lampinen et al., 2025; Nagy et al., 2025), and for forming a coherent, temporally extended notion of a self (a ‘center of narrative gravity’; Dennett, 2014).

Here we show, across two games and three experiments, that strong and robust memory distortions are observed for memory of decisions made minutes ago, even when memory accuracy is incentivised. Analysis of memory errors further reveals that, when asked what they did, people sometimes report what they believe they would have done, or what they would do now.

Results

All code, anonymized data, pre-registrations, and live demos of the experiments can be accessed at github.com/self-model/memory_reconstruction_cogsci26.

Experiment 1: people forget their wrong guesses

330 English-speaking participants (recruited via Prolific.com) completed Exp. 1 and passed two pre-registered attention checks. Due to a human error, this is more than our pre-registered *N* of 300. In line with our pre-registration (OSF.IO/3MPB4), we only analysed the first 150 participants who completed the experiment in each group. The experiment was approximately six and a half minutes long (median = 6.27), including consent forms and debrief.

In part 1 of the experiment, participants attempted to reveal all black squares in a 6-by-6 grid while touching as few white squares as possible. Each game round started with three or four black squares revealed, and participants were told that all black squares were clustered in groups of 2–4 around the revealed squares. Participants lost a point for each white square they touched, and the game round ended when all black squares were revealed. Crucially, while round 1 was identical for all participants (top left grid in Figure 1), game boards in rounds 2–4 were importantly different. In the *Symmetric* group, these boards were symmetric along the vertical axis,

In G. Boleda, I. Dautriche, M. Maldonado, C. Saldana and T. Torrent (Eds.), *Proceedings of the 48th Annual Conference of the Cognitive Science Society*

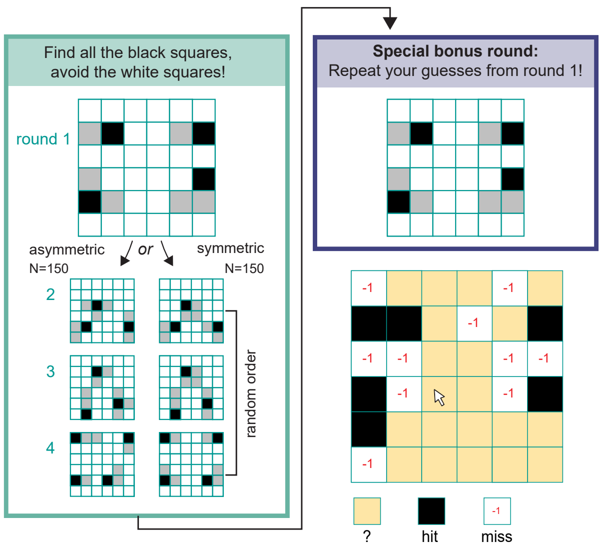

Figure 1: **Experiment 1 design**. Left: in part 1 of the experiment, participants attempted to reveal all black squares in a 6x6 grid while touching as few white squares as possible. Each round started with several black squares (one from each group) revealed (marked in black), and the others hidden (marked in gray in the figure only). Round 1 was identical for all participants, but the locations of black squares in rounds 2-4 (black and gray in the figure) were symmetric or asymmetric for different participants. Top right: In the special bonus round, participants were instructed to repeat their guesses from part 1. Bottom right: the screen as it appeared to participants during the game. Unrevealed squares were marked in yellow, and became white or black once clicked.

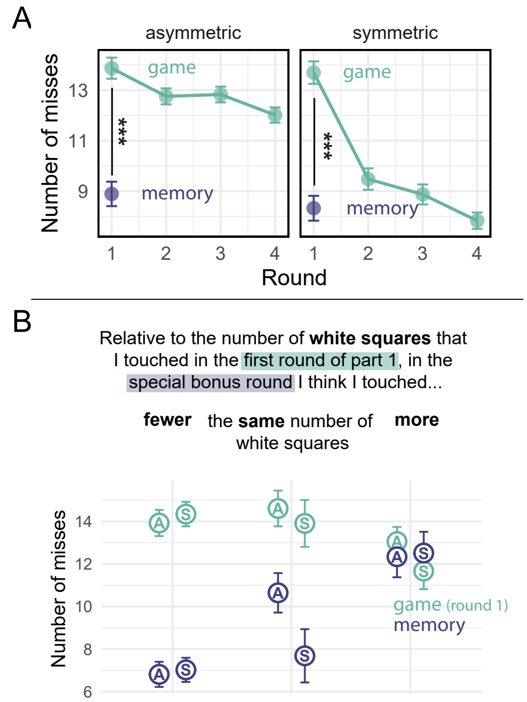

Figure 2: **Experiment 1 results**. Number of white squares touched (misses) in the game and memory rounds as a function of round number (A), and participants' answer to the debrief question (B). ***: $p < .001$. A and S correspond to the Asymmetric and Symmetric groups, respectively. Here and in all figures, error bars show the standard error of the mean.

but they were asymmetric for participants in the *Asymmetric* group. This way, by the end of the first part of the experiment, only participants in the Symmetric group learned a hidden rule: all game boards are symmetric. This allowed us to test whether this later-acquired knowledge influences recall of earlier decisions.

Once participants completed the first part of the experiment, they were introduced to a "*Special bonus round*". They were told that they would play round 1 again, but this time they would be shown the location of all hidden black squares ahead of playing. We then explained that their goal in this round is different: they need to make the same guesses they made in round 1, in the same order. We offered a \$2 bonus to participants who succeeded in doing that, and \$1 if they got "close enough". Participants continued to the bonus round only after correctly answering a comprehension check, ensuring that they understood that their goal was to repeat their guesses from round 1, not to reveal the black squares with as few guesses as before.

In round 1, the Symmetric and Asymmetric groups had a similar number of misses (Symmetric: 13.69; Asymmetric: 13.87; $t(296.77) = 0.28$, $p = .777$), but, in line with our hypothesis, the Symmetric group performed better in round 2 (9.48 vs. 12.75), 3 (8.87 vs. 12.83), and 4 (7.83 vs. 12.01; all p's $< .001$), indicating that they learned and benefited from the symmetry rule (see Figure 2A, green lines).

Our central hypothesis was that, if current knowledge distorts memory of actions performed in a different state of knowledge, the Symmetric group should remember having touched fewer white squares in round 1 compared to the Asymmetric group. Once participants know that all boards are symmetric, they may behave as if they knew it all along and misremember having guessed according to this state of knowledge from the very first board. However, this was not the case: participants in the Symmetric and Asymmetric groups had a similar number of misses in their reconstructed games (Symmetric: 8.33 misses on average; Asymmetric: 8.89; $t(297.91) = 0.82$, $p = .412$; Figure 2A, purple markers).

Crucially, however, this was not due to participants' success in remembering their number of misses: participants touched a much lower number of white squares in the memory round (mean: 8.61) compared to the first game round (mean: 13.78; $t(299) = 13.47$, $p < .001$). This underestimation effect was present in both groups of participants (Symmetric: $t(149) = 9.89$, $p < .001$; Asymmetric: $t(149) = 9.14$, $p < .001$), and was not driven by the 21 participants who recalled making no misses at all (with these participants excluded: $t(278) = 12.16$, $p < .001$). This startling underestimation of errors was also not driven by participants' attempt to finish the task quickly: a separate group of 49 participants gave a numeric estimate of how many white squares they believed they touched in round 1 before moving on to the special bonus round. This group reported making an average of 7.29 misses (SE=0.54), compared to an actual 12.49 misses (SE=0.68) in round 1 of the game. Of note, while the entire experiment took around six and a half minutes to complete, a median of only 2:13 minutes had passed between the end of round 1 and the beginning of the memory round (SE=3.52 seconds). And still, despite this short interval, participants underestimated the number of misses by more than one third.

To measure participants' awareness of this effect, we asked them whether the number of white squares they touched in the memory round was less, the same, or more than the number touched in the first round of part 1 (Figure 2B). As a group, participants knew that they underestimated the number of misses in part 1: a majority of 60% of all participants reported making fewer misses in part 2, followed by 23% for "more" and 18% for "same" responses. These counts were similar between the Symmetric and Asymmetric groups ($\chi^2(2) = 2.26$, $p = .323$).

Participants' responses were calibrated relative to their actual performance in the memory round: those who reported making fewer misses made 6.92 misses on average in the memory round, followed by 9.42 for "same" responders and 12.43 for "more" responders. While calibrated, this measure also revealed that participants truly underestimated the number of white squares they touched when playing. Participants who reported making the exact same number of misses in fact touched 4.91 fewer white squares in the memory round ($t(52) = 6.25$, $p < .001$), and those who reported committing more misses in the memory relative to the game round had a similar number of misses in the two rounds ($t(67) = -0.06$, $p = .952$; Figure 2B). Interestingly, among participants who reported making the same number of misses in the game and memory rounds, those in the Symmetric group recalled making fewer misses than those in the Asymmetric group (7.68 vs 10.65), though this effect was only marginally significant ($t(41.61) = 1.90$, $p = .064$). Together, it seems that participants tended to remember their performance as better than it actually was, and they had only partial insight into this tendency.

Experiment 2: learning distorts memory for own decisions

An underestimation of the number of misses in Exp. 1 is consistent with selective forgetting of errors. However, it is also consistent with a more general tendency to forget rather than falsely recall actions, because good performance in the game meant taking fewer actions overall. In Experiments 2 and 3 we sought to control for such effects by requiring an equal number of recalled actions for good and bad performance in the game. To do this, we designed a new game.

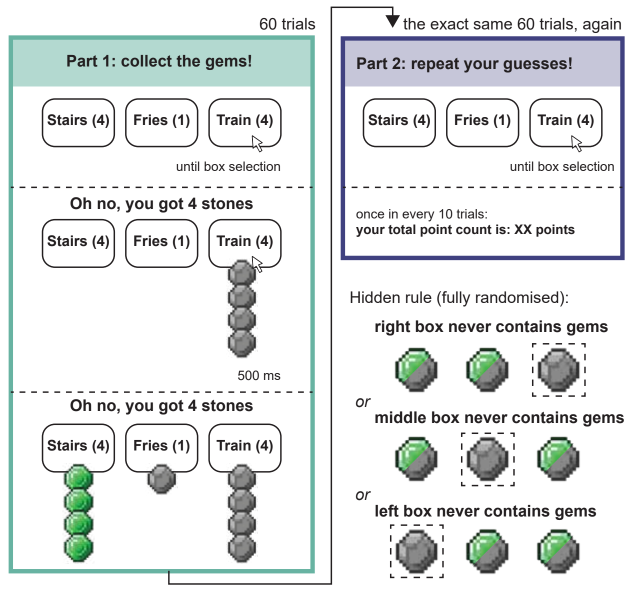

Figure 3: **Experiment 2 design**. Left: in part 1 of the experiment, participants attempted to collect as many gems as possible by choosing between boxes. Each box was labelled with a random noun and a number, indicating the number of stones or gems in it. Right: in part 2, participants attempted to repeat their guesses from part 1. One of the three positions (the *stones only* position — left, middle, or right, for different participants) never included gems.

In Experiment 2, a total of 117 English-speaking participants were recruited via Prolific.com. 100 participants passed our pre-registered inclusion criteria (pre-registration: OSF.IO/K8AT9) and were included in the final sample. The experiment consisted of two parts (see Figure 3), each with 60 trials, and lasted approximately 11 minutes (median = 10.7). In part 1, participants attempted to accrue points by revealing hidden gems and in part 2 they attempted to repeat their guess sequence from part 1.

On each trial in part 1, participants were presented with three boxes, each labelled with a unique noun (sampled without replacement from a list of common nouns) and a number (sampled from a uniform distribution over one to five). Participants were told that the number indicated the number of items in the box: stones or gems. Out of the three boxes, only one contained gems (worth one point each), and two contained worthless stones. Participants' task was to accumulate as many points as possible by collecting gems. After making their decision, participants received feedback: first about the content of their chosen box and then about the content of all three boxes. A counter on the top right of the screen indicated the points participants had accumulated. As an attention check, participants were told that boxes labelled with fruit words would always contain gems. Only three trials included fruit words, and participants had to choose the fruit word in at least two out of these catch trials to be included. Catch trials were excluded from the analysis (as pre-registered). We designed Exp. 2 to induce a non-trivial decision-making process per trial: reading the words, considering the numbers, and perhaps entertaining different patterns or rules that could govern the location of the gems. Indeed, gem locations followed a hidden rule: gems could only appear in two of the three positions. The location of the *stones only* position – left, middle, or right – was randomised for each participant (Figure 3, bottom right). This allowed us to test how learning affects memory for past decisions.

Immediately after completing part 1, participants were introduced to part 2, where they were presented with the exact same sequence of 60 box arrangements as before. However, their task was now to repeat the same sequence of guesses as in part 1. Participants were awarded 3 points for each accurate reconstruction. Feedback about the total accumulated points was delivered once every 10 trials, but no trial-wise feedback was provided.

In part 2, participants accurately recalled their choices in part 1 on 66% of trials, and significantly higher than chance (1/3; $t(99) = 25.24$, $p < .001$). In line with an outcome bias in memory for actions, participants' ability to reconstruct their past decisions was significantly higher when they were correct in part 1 (70%) compared to when they were incorrect in part 1 (63%; $t(99) = 6.43$, $p < .001$; see Figure 4A). That is, participants remembered their successes better than their failures.

Participants learned to avoid the *stones only* position as they played part 1 of the game ($P(\text{stones only}) = 0.25$, significantly lower than 1/3; $t(99) = -7.51$, $p < .001$), and the probability of choosing the *stones only* position decreased as a function of trial number (a Wilcoxon test on subject-specific slopes from a logistic regression, due to outliers in the sample: $V = 1,118.00$, $p < .001$; see Figure 4B, green line). 22% of participants explicitly mentioned noticing a regularity in the location of gems when asked at the end of the experiment, but learning was observed among other participants as well ($P(\text{stones only} \mid \text{no rule mentioned}) = 0.28$, significantly lower than 1/3; $t(77) = -5.32$, $p < .001$). This allowed us to test whether the detection of a general rule across all trials distorts people's memory of their past actions in individual trials.

In line with a "knew-it-all-along" effect (Fischhoff, 1975; Wood, 1978), the probability of choosing the *stones only* position in part 2 was 0.24, significantly lower than chance ($t(99) = -7.85$, $p < .001$), and, crucially, significantly lower than in part 1 ($t(99) = -2.20$, $p = .030$). Visual inspection of Figure 4B reveals that participants misremembered their decisions as if they avoided the *stones only* position from the very first trial: unlike in part 1, this probability was relatively stable throughout part 2, with a marginal effect of trial number (a Wilcoxon test on subject-specific slopes: $V = 1,989.00$, $p = .066$). Accordingly, the learning slope in part 1 was significantly steeper than in part 2 ($V = 1,663.00$, $p = .005$).

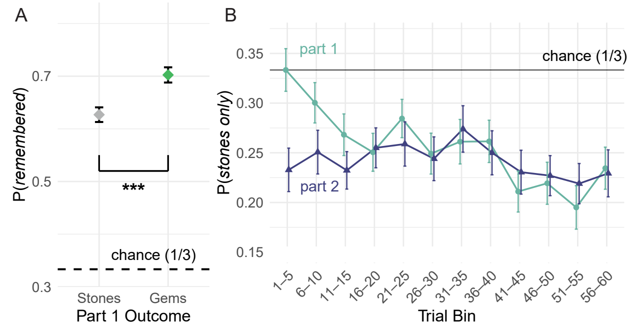

Figure 4: **Experiment 2 selective memory effects**. A: probability of accurately remembering a decision as a function of outcome in part 1. B: probability of choosing the *stones only* position as a function of trial number in parts 1 and 2.

Experiment 3: memory distortions are not strengthened by explicit rule knowledge

Participants in Experiment 2 reconstructed their guesses as if they knew the hidden rule all along, and were better able to remember their correct guesses than their incorrect ones. Experiment 3 (see Figure 5) sought to examine whether these distortions are mediated by processes that are under participants' conscious control. To this end, we directly told participants about the hidden rule before they started part 2, and revealed where the gems had been in part 1 of the experiment by highlighting the box in green. We reasoned that if participants were knowingly biased to guess the correct option more frequently, memory distortions would become stronger once the correct option is marked as such. In contrast, if these effects are driven by a confusion between action and feedback ('automatic assimilation'; Fischhoff, 1975), marking the correct option should decrease memory distortions (Jacoby & Kelley, 1987).

In Exp. 3, we also asked participants to rate their confidence on an analog scale after each decision in part 2, and reduced the number of trials in each part to 50 to keep the experiment short. A total of 113 English-speaking participants were recruited via Prolific.com. 100 participants passed our pre-registered inclusion criteria (pre-registration: OSF.IO/RDBS7) and were included in the final sample. The experiment was approximately 13 minutes long (median = 13.2).

We replicated previous results from Experiment 2. In part 1, participants learned to avoid the *stones only* position, selecting it with probability 0.28, significantly below chance (1/3; $t(99) = -5.33$, $p < .001$). A Wilcoxon test on subject-specific learning slopes from a logistic regression showed that participants improved as a function of trial number ($V = 1,339.00$, $p < .001$). 27% of participants explicitly noticed the hidden rule, though participants who did not notice the rule also showed signs of learning ($P(\text{stones only} \mid \text{no rule mentioned}) = 0.31$, significantly lower than 1/3;

$t(72) = -2.10$, $p = .039$).

In part 2, participants recalled their choices with an accuracy of 65%, significantly higher than chance (1/3; $t(99) = 22.09$, $p < .001$). Similar to Exp. 2, memory was better for trials with correct decisions (69%) compared to incorrect decisions (62%; $t(99) = 2.90$, $p = .005$, see Figure 5B). This effect was not significantly different from the one observed in Exp. 2 ($t(152.11) = 0.50$, $p = .617$). Again, participants chose the *stones only* position significantly below chance (0.26; $t(99) = -5.05$, $p < .001$). Unlike in Exp. 2, however, this probability was not significantly different from the probability of choosing the *stones only* position in part 1 ($t(99) = 1.41$, $p = .160$). This apparent lack of difference was also reflected in participants' reconstructed learning slopes: unlike in Experiment 2, in Experiment 3 learning slopes were significantly negative in part 2 ($V = 1,475.00$, $p < .001$), with no significant difference between the learning slopes of part 1 and 2 (a Wilcoxon test on the difference in subject-specific slopes across parts: $V = 2,670.00$, $p = .497$; see Figure 5D). The difference between the two experiments is in line with an unconscious origin of outcome-based memory distortions: once information about the location of gems was made available during retrieval, participants could metacognitively "undo" its biasing effects on the experience of familiarity, even if not fully (Jacoby & Kelley, 1987; Jacoby et al., 1989).

Finally, confidence ratings revealed that participants had metacognitive awareness of the accuracy of their memory judgments, with an overall higher confidence in part 2 when correctly recalling their decisions from part 1 (mean confidence = 60.78, SE = 2.20) versus when incorrectly recalling their decisions (mean = 47.44, SE = 2.20; $t(99) = 11.31$, $p < .001$; see Figure 5C). Critically, participants were also more confident when correctly recalling that they revealed gems in part 1 (mean = 64.09, SE = 2.14) compared to when correctly recalling that they revealed stones in part 1 (mean = 58.99, SE = 2.29; $t(99) = 4.62$, $p < .001$). That is, participants were not only biased to report having been correct in part 1 of the experiment; they were also more confident when doing so.

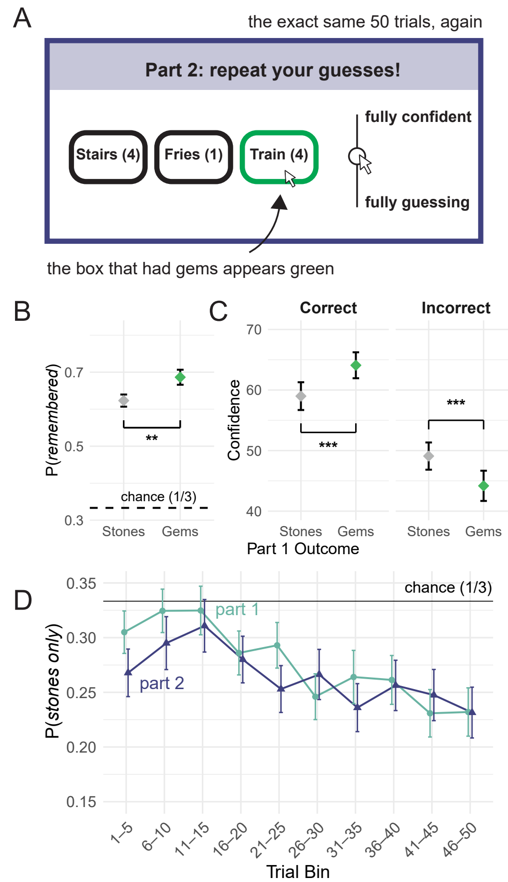

Figure 5: **Experiment 3 methods and results**. A: Unlike in Exp. 2, in part 2 of Exp. 3 participants were shown the true location of gems in part 1, and rated their confidence after each decision. B: probability of accurately remembering a decision as a function of outcome in part 1. C: confidence as a function of outcome in part 1 and memory accuracy. D: probability of choosing the *stones only* position as a function of trial number in parts 1 and 2.

Analysis of memory errors reveals effects of model-based reconstruction

So far, we mostly focused on the effects of outcome and rule knowledge on the accuracy of memory for past decisions, showing that participants better remember their successful guesses than their unsuccessful ones. In the following, we zero in on the effects of outcome and rule knowledge not on the proportion of errors, but on the types of errors participants make.

As a first test we focused on a subset of trials in which participants failed to remember that they had made a correct choice (that is, chose the box containing gems) in part 1. In these trials, participants had chosen the box of gems in part 1 and in part 2 either selected the *stones only* position or the position that sometimes (but not in this trial) contained gems (see Figure 6A). This test deconfounds the memory bias against the *stones only* position from a memory bias in favour of gems or from a memory for originally biased actions. Given that different participants had different numbers of error trials, we fitted a mixed effects logistic regression model with a random intercept for participant. In Experiment 2, the estimated probability of selecting the *stones only* position in the specified subset of trials was 0.4, and significantly below the chance level of 0.5 ($\beta = -0.39$, SE = 0.11, $z = -3.60$, $p < .001$). This was replicated in Experiment 3, where the estimated probability of selecting the *stones only* position was 0.38 ($\beta = -0.51$, SE = 0.13, $z = -3.94$, $p < .001$). These results show a robust and systematic avoidance of the *stones only* position in errors controlling for reward, which we interpret to be a bias rooted in (faulty) model-based memory reconstruction: participants avoided choosing the option which they believed they would not have chosen, given what they currently know (but did not

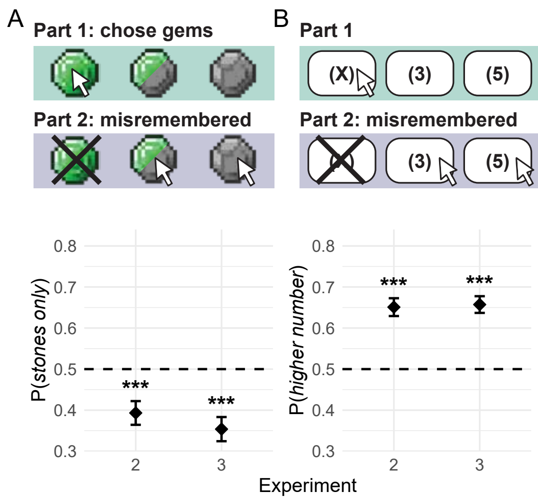

Figure 6: **Experiment 2 and 3 error analysis**. Left: participants were significantly less likely than chance to choose the *stones only* position, conditioned on having forgotten choosing the gems option in part 1. Right: Conditioned on having forgotten their decision in part 1, participants were significantly more likely to choose the option that was associated with a higher number between the two remaining options.

necessarily know then).

A second way in which participants' responses in part 2 could be shaped by beliefs about what they were likely to do is through the effect of numbers on choice. In part 1, chosen boxes were associated with numbers significantly above the mean value of 3 (Exp. 2: mean = 3.27, $t(99) = 11.37$, $p < .001$, Exp. 3: mean = 3.45, $t(99) = 11.37$, $p < .001$). This behaviour is rational: box labels indicated the number of gems or stones in the box, and while gems were associated with a reward, stones were not associated with any cost.

Conditioning on participants not remembering their response in part 1, and asking which of the remaining options was chosen (Figure 6B), participants were more likely to choose the option that was associated with a higher number in both Exp. 2 (P(higher number) = 0.65, $\beta = 0.63$, SE = 0.10, $z = 6.43$, $p < .001$) and Exp. 3 (P(higher number) = 0.66, $\beta = 0.65$, SE = 0.10, $z = 6.85$, $p < .001$). Again, memory errors for self-actions were not random: they tracked the decisions that would make sense, suggestive of a model-based reconstruction.

Discussion

Previous research has identified biases in how people remember their decisions and actions. For example, participants are systematically biased to misremember their answers to questions in an IQ test completed several months earlier as more accurate than they were, to report never having seen questions if they did not solve them correctly initially, and to recall having correctly answered questions that they had in fact never encountered (Chew et al., 2020). Similar "knew it all along" effects are observed in people's memory of their past predictions, for example of election results (Blank et al., 2003; Calvillo & Rutchick, 2014; Chen et al., 2019; Powell, 1988) and Super Bowl outcomes (Bonds-raacke et al., 2001).

Here, we show that the tendency to selectively misremember one's decisions as more accurate than they were can be observed even for decisions made less than three minutes ago, without any intervening task (*cf.* Arnold & Lindsay, 2007), and with adult participants (*cf.* Gopnik & Astington, 1988). Error patterns reveal that these biases are not solely due to better encoding of successful decisions (an effect that would be in line with a dopamine-driven memory boost; Wittmann et al., 2005): focusing on error trials only, we show that participants reconstruct their past decisions based on what they believe would have made sense to do ("*I would not have chosen the left position, it always contains stones*"; "*I would have chosen the box with the higher number*"). This is an instance of model-based reconstruction: using a schema, or a model, to fill in missing details from a patchy memory record. Indeed, biased reconstruction of this kind has been proposed to play a key role in hindsight bias (Müller & Stahlberg, 2007; Schwarz & Stahlberg, 2003). Using a gamified task, we show that biased reconstruction distorts memory of actions taken prior to learning not only individual facts, but also general patterns (here, the location of the *stones only* position).

Using the current decision policy as a schema for past decisions One possible mechanism that could explain these memory biases is reconstruction of past decisions based on a readily available self-model: one's current decision policy. In order to estimate what they would have decided, people may simply decide again. This strategy could be efficient: assuming that we mostly need to recall recent decisions, and that decision policies only change slowly, the current decision policy is a close replica of the policy that was used to make the recalled decision. Furthermore, using the decision policy as a memory schema does not require agents to maintain multiple "time-stamped models" of their past selves.

Crucially, this heuristic would break exactly when attempting to recall decisions that were made under a different decision policy, for example before learning about a regularity in the environment, as in our task. According to this account, participants' better ability to recall their early mistakes in Exp. 3 compared to Exp. 2 (compare Figure 5D and Figure 4B) may have been exactly because, by revealing the location of the gems in part 2 of Exp. 3, we made it impossible for them to rely on their current preferred option as a guide for the option that seemed best when playing part 1, when the gems were hidden.

A bias to recall past decisions as consistent with one's current behavioural policy may be beneficial, for example, by allowing learners to train their habitual system through replaying not the decisions that were made, but the decisions that should have been made. This, however, may come at the cost of an inaccurate impression of a temporally stable, unchanging self.

Acknowledgements

This work was supported by the British Academy/Leverhulme Small Research Grants scheme [SRG2425\252032], awarded to MM. MM is supported by a postdoctoral research fellowship at All Souls College, Oxford. TG was supported by grants from the Stanford Institute for Human-Centered Artificial Intelligence (HAI) and from the Cooperative AI Foundation.

References

Arnold, M. M., & Lindsay, D. S. (2007). "I remember/know/guess that I knew it all along!": Subjective experience versus objective measures of the knew-it-all-along effect. *Memory & Cognition*, *35*(8), 1854–1868.

Bartlett, S. F. C. (1995, June). *Remembering: A Study in Experimental and Social Psychology* [Google-Books-ID: WG5ZcHGTrm4C]. Cambridge University Press.

Biderman, N., Bakkour, A., & Shohamy, D. (2020). What Are Memories For? The Hippocampus Bridges Past Experience with Future Decisions. *Trends in Cognitive Sciences*, *24*(7), 542–556.

Blank, H., Fischer, V., & Erdfelder, E. (2003). Hindsight bias in political elections. *Memory*, *11*(4-5), 491–504.

Bonds-raacke, J. M., Fryer, L. S., Nicks, S. D., & Durr, R. T. (2001). Hindsight Bias Demonstrated in the Prediction of a Sporting Event. *The Journal of Social Psychology*, *141*(3), 349–352.

Calvillo, D. P., & Rutchick, A. M. (2014). Political knowledge reduces hindsight memory distortion in election judgements. *Journal of Cognitive Psychology*, *26*(2), 213–220.

Chen, Y.-H., Cheng, H.-P., Lu, Y.-W., Lee, P.-H., Northoff, G., & Yen, N.-S. (2019). Can knowledge of election results change recall of our predictions? Neural correlates of political hindsight bias (M. E. Le Pelley, Ed.). *PLOS ONE*, *14*(10), e0220690.

Chew, S. H., Huang, W., & Zhao, X. (2020). Motivated False Memory. *Journal of Political Economy*, *128*(10), 3913–3939.

Dennett, D. C. (2014). The self as a center of narrative gravity. In *Self and consciousness* (pp. 103–115). Psychology Press.

Fischhoff, B. (1975). Hindsight is not equal to foresight: The effect of outcome knowledge on judgment under uncertainty. *Journal of Experimental Psychology: Human Perception and Performance*, *1*(3), 288–299.

Gershman, S. J., & Daw, N. D. (2017). Reinforcement Learning and Episodic Memory in Humans and Animals: An Integrative Framework. *Annual Review of Psychology*, *68*(Volume 68, 2017), 101–128.

Gopnik, A., & Astington, J. W. (1988). Children's Understanding of Representational Change and Its Relation to the Understanding of False Belief and the Appearance-Reality Distinction. *Child Development*, *59*(1), 26–37.

Hemmer, P., & Steyvers, M. (2009). A Bayesian Account of Reconstructive Memory. *Topics in Cognitive Science*, *1*(1), 189–202.

Jacoby, L. L., & Kelley, C. M. (1987). Unconscious Influences of Memory for a Prior Event. *Personality and Social Psychology Bulletin*, *13*(3), 314–336.

Jacoby, L. L., Woloshyn, V., & Kelley, C. (1989). Becoming famous without being recognized: Unconscious influences of memory produced by dividing attention. *Journal of Experimental Psychology: General*, *118*(2), 115–125.

Lampinen, A. K., Engelcke, M., Li, Y., Chaudhry, A., & McClelland, J. L. (2025, December). Latent learning: Episodic memory complements parametric learning by enabling flexible reuse of experiences [arXiv:2509.16189 [cs]].

Levin, M. (2024, April). Self-improvising Memory: A perspective on memories as agential, dynamically-reinterpreting cognitive glue.

Mazor, M., Firestone, C., & Phillips, I. (2026). Pretending not to know reveals a capacity for model-based self-simulation. *Psychological Science*, 09567976251409747.

Mazor, M., Seghezzi, S., & Manohar, S. (2026). Remembering what you did: Episodic memory for self-actions. *Neuroscience and Biobehavioral Reviews*.

Müller, P. A., & Stahlberg, D. (2007). The Role of Surprise in Hindsight Bias: A Metacognitive Model of Reduced and Reversed Hindsight Bias. *Social Cognition*, *25*(1), 165–184.

Nagy, D. G., Orbán, G., & Wu, C. M. (2025). Adaptive compression as a unifying framework for episodic and semantic memory. *Nature Reviews Psychology*, *4*(7), 484–498.

Powell, J. L. (1988). A Test of the Knew-It-All-Along Effect in the 1984 Presidential and Statewide Elections. *Journal of Applied Social Psychology*, *18*(9), 760–773.

Schwarz, S., & Stahlberg, D. (2003). Strength of hindsight bias as a consequence of meta-cognitions. *Memory*, *11*(4-5), 395–410.

Spens, E., & Burgess, N. (2024). A generative model of memory construction and consolidation. *Nature Human Behaviour*, *8*(3), 526–543.

Wittmann, B. C., Schott, B. H., Guderian, S., Frey, J. U., Heinze, H.-J., & Düzel, E. (2005). Reward-Related fMRI Activation of Dopaminergic Midbrain Is Associated with Enhanced Hippocampus- Dependent Long-Term Memory Formation. *Neuron*, *45*(3), 459–467.

Wood, G. (1978). The knew-it-all-along effect. *Journal of Experimental Psychology: Human Perception and Performance*, *4*(2), 345–353.