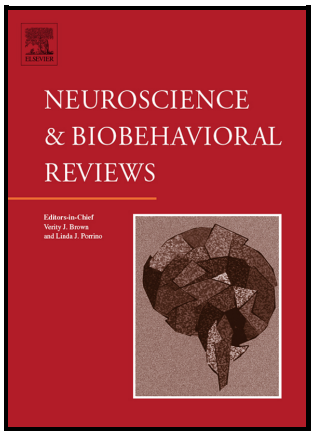
# Remembering what you did: episodic memory for self-actions

Matan Mazor, Silvia Seghezzi, Sanjay Manohar




PII: S0149-7634(26)00182-X

DOI: https://doi.org/10.1016/j.neubiorev.2026.106725

Reference: NBR106725

To appear in: *Neuroscience and Biobehavioral Reviews*

Received date: 4 March 2026
Revised date: 27 April 2026
Accepted date: 30 April 2026

Please cite this article as: Matan Mazor, Silvia Seghezzi and Sanjay Manohar, Remembering what you did: episodic memory for self-actions, *Neuroscience and Biobehavioral Reviews,* (2026) doi:https://doi.org/10.1016/j.neubiorev.2026.106725

# Remembering what you did: episodic memory for self-actions

Matan Mazor[1], Silvia Seghezzi[2] & Sanjay Manohar[1]

[1]University of Oxford
[2]Birkbeck, University of London



Correspondence concerning this article should be addressed to Matan Mazor, All Souls College, High Street, Oxford OX1 4AL. E-mail: matan.mazor@all-souls.ox.ac.uk

**Abstract**

Episodic memory stores not only passively experienced events, but also one's own past actions and decisions. Despite their critical role for learning about the world and about the self, little is known about how such memories for self-actions are stored and retrieved. We argue that memory for self-actions plays three key roles in the cognitive economy: it scaffolds memory for environmental events, enables learning from delayed feedback, and allows individuals to learn about their own abilities and preferences. A synthesis of evidence from behaviour and psychopathology is consistent with an organizing framework: memory for self-actions draws on a generative self-model — a simplified schema of one's own cognition. This framework helps explain why memory for habitual actions is particularly vulnerable to memory distortions, manifesting as confabulations in amnesic patients and obsessive doubt in individuals with OCD. We further propose that an accurate self-model may partially compensate for hippocampal memory loss in Alzheimer's disease. We suggest that much can be learned from studying the cognitive and neural mechanisms underlying the ability to remember one's own decisions and actions, and identify critical questions for advancing our understanding of this important but neglected aspect of human memory.

Keywords:
episodic memory, self-model, metacognition, confabulation, OCD, Alzheimer's disease

Episodic memory is, put simply, a mental record of your past experiences. Many of the episodes that are recorded in episodic memory feature the self not only as an experiencer, but as an active agent too. This is clearly revealed when reflecting on everyday failures of episodic memory: “where did I put the jacket?”, “did I already add salt to the salad?”, “did I tell him this anecdote before?”, “did I remember to turn off the stove before leaving?”, “did I take my pills this morning?”.

Interestingly, however, when studied in the lab and measured in the clinic, episodic memory is most often operationalised as memory for facts about the external world. A typical episodic memory test involves a learning phase in which the participant is exposed to information, followed by a recognition or a recall phase. The learned information varies by study: a recent meta-analysis of episodic memory [1] includes studies in which participants’ memory was measured for different categories of stimuli: verbal stimuli such as words and sentences (371 studies), images (189 studies), movie clips (23 studies), locations of objects (67 studies), routes through space (24 studies), faces (59 studies), odors, tastes and colors (9 studies). Tellingly, none of the 617 studies surveyed in this analysis measures participants’ ability to remember what they did or decided.

The dominance of passive memory paradigms over more naturalistic paradigms makes sense from an experimental perspective, especially when the goal is to isolate mechanisms of encoding and retrieval. Above-chance memory of the random words presented in a previous learning phase is solid evidence of successful encoding and retrieval. In contrast, a participant performing above-chance in indicating which decisions they made in a previous game (for example, the sequence of turns they took in navigating a maze), may be manifesting their successful encoding or retrieval (“I remember I turned right here”), but it may also reflect an ability to re-enact the same decisions upon demand (“I would have turned right here”). Thus, when tested, memory for self-actions is inherently contaminated with our ability to estimate how we would behave in different situations (model-based self-simulation), or our consistency in behaving similarly when presented with the same setting again. For this reason, empirical studies of memory for actions typically instruct participants to perform arbitrary actions [2–5], making it impossible for participants to rely on their beliefs about what they would or would not have done (see Box 1).

Importantly, the very same considerations that make memory for voluntary actions a poor measure for information storage and retrieval also highlight an important fact: episodic memory may rely on radically different mechanisms when information about the world is entangled with information about one’s own decisions. When studied in the lab, episodic memory is neatly separated into encoding and retrieval phases, but in real life it is embedded in a continuous loop of information encoding, decision-making, and the reconstruction of these past decisions. This discrepancy between in-lab and real-life episodic memory may be one reason that performance in lab-controlled tasks is a poor predictor of memory for life events [6]. To understand episodic memory we should therefore understand how people remember their own actions — be it through direct retrieval or reconstruction.

## Three unique benefits of memory for self-actions

Despite not being systematically studied, an ability to remember one's own past decisions and actions has substantial benefits. Let us name three. First, given how intimately entangled experiences are with self-decisions and self-actions, a record of one's past decisions, or an ability to reliably reproduce past decisions via simulation, may provide a *scaffold* for episodic memory more generally. If agents can reliably infer their past actions from noisy, or even missing, mnemonic evidence, inferred self-actions can become cues for elaborating and enriching memories about aspects of the environment too. As an example, I may remember that I cycled home through the city last Thursday, and that I normally prefer the river path unless it is flooded. Together, I can infer that the river path must have been flooded last Thursday: a fact about the external environment. A scaffolding role for self-decisions may also explain some of the superiority of "active learning"—learning regimes in which exposure to new information is interleaved with and mediated by students' decisions and actions—over traditional, passive learning [7–9].

Second, remembering one's past decisions may be crucial for learning from one's mistakes and successes and for making better decisions in the future. Some models of decision-making assume that agents have access to a record of their past decisions, which they consult when coming to make new decisions by computing the expected value of different options "on the fly" ["Episodic Reinforcement Learning" 10,"Case-Based Decision-Theory" 11]. Furthermore, when actions and action-outcomes are separated in time, credit assignment in reinforcement learning depends on having a record of the actions that led to the observed outcome. It has recently been suggested that, for this reason, motor learning critically depends on motor working memory: a short-term record of one's movements [12–14]. Episodic memory for self-actions may well serve a similar function over longer time scales, focusing on decisions rather than motor content. Indeed, an involvement of episodic memory in decision-making is supported by a functional involvement of the hippocampus in decision-making tasks, and by observations of deficits in decision-making among hippocampal patients [e.g., 15, for a review, see 16].

Some of the decision-making benefits of episodic memory can in principle be achieved by relying on memory for the value of states rather than memory for the decisions that led to these states (for example, remembering that the a certain dish in the menu is tasty rather than remembering my decision to order it a second time). Importantly, however, memory for affective or valenced states is unreliable and subject to biases: the remembered intensity of emotions fades with time [17–19], memory for emotions is affected by post-event knowledge [20,21] and by current appraisals [22], and memory for affective episodes neglects to factor in episode duration, focusing on the peak and the end of the emotional experience [23]. Memory for our past actions can compensate for these biases and limitations. I may not remember exactly how bored I felt when watching a certain movie at the cinema, but I do remember staring at the ceiling at some point – an objective, measurable signature of my boredom at the time. Later, I can infer the subjective value associated with the film based on my observed reactions to it [a form of "representational exchange", 24]. This way, memory for our past actions and decisions

can allow us to estimate the subjective value associated with different states, in ways that may be more reliable than memory for the subjective value itself as experienced at the time.

## The self-model and memory for actions

Episodic memory is reconstructive: while some details appear on record, other details are omitted and need to be inferred [25–27]. This process can be described as an approximate a Bayesian inference: noisy, partial retrieved information about a past episode is combined with prior knowledge to "invert the forward memory model" [that is, the internal model that describes how information is retained in memory 28,29] and produce a best guess regarding the world state most likely to have triggered the retrieved memory trace. This prior knowledge, which may be represented implicitly, in non-declarative form, can be described as a *schema* or a *model*. Model-based reconstruction leaves traces on behaviour in the form of *schema-based distortions*: things are often remembered as more generic, or similar to a prototype or a schema, than they actually were. For example, people are biased to remember individual fruits as being more similar in size to a prototypical category member [28].

Agents may consult domain-specific models or schemas when reconstructing different aspects of their memory, and the usefulness of a model can be measured as its efficiency in compressing behaviourally-relevant information, that is, in reliably representing more information given memory storage constraints. For example, chess players have improved memory for chess boards, presumably because familiarity with the rules makes for more efficient representations of game boards [30,31]. As mentioned earlier, one such system that needs to be modelled is the agent itself: if I know what I would do in different circumstances, I can accurately reconstruct my actions even when they are not encoded explicitly. Since the self is the main character of our episodic memories, many of our memories feature decisions and actions taken by us. As a result, having a good model of the self becomes a highly effective compression device: it allows agents to represent more information using less storage space.

Indeed, it has been suggested that the self, or the self-model, serves as a "superordinate schema" [32] for encoding information in memory, which may explain why information is better remembered when encoding involves a comparison to the self [33]. Critically, for this model to be accurate it needs to be maintained and updated by the agent in light of new evidence: information about our own actions and decisions. While some knowledge about our motivations and preferences may be available through introspection, the ability to directly introspect over private states is limited [34–36,37; see Box 2]. We therefore need to learn about our motivations and preferences from an internal record of our past decisions and actions. As a result, the relationship between memory for own actions and the self-model is a reciprocal one: the model is used for reconstruction of past actions, and it draws on past actions to maintain and update its content. This circularity means that biases in the specification of the self-model may distort how individuals remember their past decisions and actions, which in turn perpetuates these very biases in how they model their motivations and preferences (see Box 3).

As an example of model-based reconstruction of memory for self actions, consider the experience of looking for your lost keys. Past you has put them somewhere, and present you is now trying to recover where. You may have a vague memory of seeing the keys in the bedroom, but your self-model tells you that you are unlikely to have left your keys there (see Fig. 1): you are much more likely to have left them in the kitchen, in your office, or by the entrance door. This prior distribution over locations does not need to be explicitly represented for it to be accessible: it can be dynamically derived from a generative, probabilistic self-model via simulation [38,39]. The memory trace (the likelihood) can then be combined with the model-derived prior probability of leaving the keys in different places to produce a posterior distribution over locations: a model-based memory reconstruction. The self-model can therefore guide the key search ("I would not have left my keys here"), and, crucially, be informed and updated upon eventually finding the keys ("It turns out that I sometimes leave my keys in my coat").

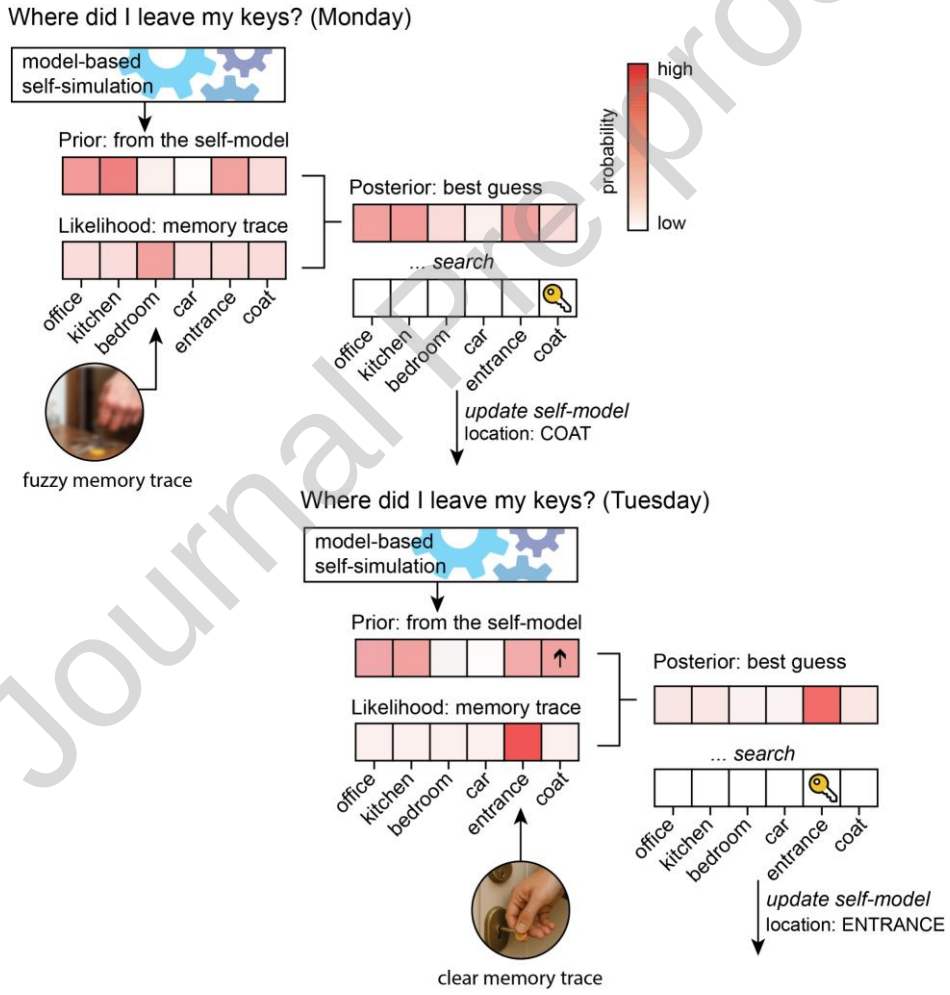



*Fig. 1: the self-model in memory for self-actions*. On Monday (top left), a person is looking for their keys. They use the self-model to simulate possible locations in which they may have left their keys, producing a prior distribution in which the kitchen is most likely, followed by the office and the entrance door. This distribution is integrated with a fuzzy memory trace of leaving the key at the bedroom, producing a posterior distribution which guides the physical search. Upon finding the keys in the coat, the person updates their self-model: it turns out that they sometimes leave their keys in the coat. On Tuesday

(bottom half), this updated self-model is used to generate likely actions, producing a new prior distribution. A clear memory trace of having left the keys in the entrance door dominates the posterior, leading the physical search.

In this Bayesian framework, the role of a prior is fulfilled by a self-model: a simplified representation of one's cognitive machinery [a form of metacognitive knowledge 40] that can be simulated to estimate of one's most likely behaviour in a given setting (for example, where one was most likely to leave their keys). But it is possible to imagine that some of this function is covered not by a simplified model of the cognitive system, but by the cognitive system itself, by *re-enacting* the decision-making process (for example, pretending to have the keys in my hand and deciding where to put them). As a result, memory for self actions may rely on a mixture of memory traces of specific events, prior metacognitive beliefs about the self (represented, for example, in the form of a probabilistic self-model), and an ability to reliably produce the same action when put under the same conditions again.

While this computation-level description is agnostic with respect to the algorithmic and neural-instantiation details [41], clinical and neuroimaging studies of episodic memory and metacognition provide a framework to map these computations to specific brain structures and modes of implementation. Broadly, memory traces are predominantly encoded and initially stored in the hippocampus, whereas schemas and models are instantiated in the cortex [42–44]. Schemas that relate to the self [a form of metacognitive knowledge 40] are likely encoded in a distributed manner, with the ventromedial prefrontal cortex serving as a key hub [45–49]. The reconstruction process may also draw on schemas of the behaviour of agents more generally, recruiting parts of the Theory of Mind network [50], including the temporoparietal junction. To the extent that one's current behavioural policy is used as a prior for one's past decisions and actions, reconstruction of one's past decisions may also draw on the same brain regions that drive decision-making, including structured knowledge about contingencies and preferences in the orbitofrontal cortex [51–53]. Episodic memory of the motor action itself, rather than the decision to take a particular motor action, may draw more heavily on neural substrates of motor imagery [54]. Memory traces may be integrated with schemas through iterative hippocampal pattern completion and cortical reinstatement [43,55,56].

## Pathologies of memory for self-actions

In the framework we put forward here, an accurate self-model enhances memory capacity for self-actions by allowing actions and decisions to be reconstructed even in the absence of a clear memory trace. Instead of relying on the successful encoding of each decision and action to long-term memory, a generative self-model can be used to approximate what one would do under given circumstances. This selective dependence of memory for self-actions on the self-model provides a useful framework for thinking about certain pathologies. Here we will focus on three: pathological confabulation in memory loss, Alzheimer's disease, and compulsive checking symptoms in obsessive compulsive disorder.

## Confabulation as reconstruction without recall

Some cases of clinical memory loss are accompanied by a tendency to misremember imagined events as if they had actually happened in real life, that is, to confabulate. Confabulations commonly occur in patients with Korsakoff syndrome: a chronic memory disorder caused by severe deficiency of thiamine (vitamin B1), often the result of excessive alcohol consumption [57]. Confabulations may also be observed in other amnesic patients [58–60], and can result from lesions to a wide network of brain regions, including the orbitofrontal cortex and the mamillary bodies: a crucial node in the integration of hippocampal memory traces with cortically represented schemas and models [58,61]. Crucially, confabulations are almost never observed when both hippocampi are fully lesioned [62,63], suggesting that they require some hippocampal function. During healthy retrieval of memories of self-action, we propose that the hippocampus encodes an efficient, compressed representation of the actions given our self-knowledge. Reconstruction then involves the integration of this stenographic representation with model-based self-knowledge, which is presumably represented in the prefrontal cortex (Fig. 1). However, in pathological states such as Korsakoff's syndrome, when retrieval occurs, the hippocampal outputs are corrupted but a sense of familiarity is still preserved [64,65]. Reconstruction is then based on noisy inputs, resulting in memories that are in fact dominated by random samples from the generative self-model.

Indeed, the predominant form of confabulation involves personal habits: action routines that are more strongly represented in the self-model. Confabulators are more likely to falsely remember having performed a habitual action, one that is part of their normal routine, than any other form of confabulation [60,62]. "Habit confabulations" are particularly striking when seen in the hospital: hospitalized patients, when asked what they did yesterday, would often falsely report events from their life routine from before admission. A famous case in point is that of patient B.E. [59] who was seen at the neuropsychology department seven weeks after an operation to clip an aneurysm of the anterior communicating artery. B.E., a shopkeeper, used to earn extra money by performing stocktakes for other shops. Following his aneurysm, B.E. repeatedly confabulated memories of recent stocktakes that he performed, or upcoming ones that he should complete. Importantly, these false memories did not match any single event (they often involved shops he had never worked for), but fitted a general template of actions that were typical for him before the operation. This tendency to generate overly prototypical samples is expected if confabulations reflect a model-based reconstruction of noisy, or missing, mnemonic evidence, based on a generative self-model. Indeed, similar "schema-based distortions" [28] are observed when people try to enact how they would behave had they not known something [66] and when artificial generative networks attempt to generate new samples from a learned distribution [67].

## Alzheimer's disease increases reliance on the self-model

Unlike confabulation due to Korsakoff's for example, Alzheimer's disease (AD) damages the medial temporal lobe structures themselves. This may result in loss not only of the memory content, but also the sense of familiarity [68]. Semantic memory is relatively preserved in AD

[69,70], and we suggest that, similar to pathological confabulation, the self-model would be intact.

Our account predicts that when episodic memory is lost, we may still be able to retrieve elements of it using our self-model. For example when patients forget where they placed their wallet, they can use their self-model to improve access to that memory by simulating through possible places they may have placed it. In this situation, the ability to retrieve lost memories should be higher among individuals whose actions are predictable to themselves. Crucially, unlike confabulating individuals, AD patients do not normally mistake their model-based priors for specific episodes. For example, when asked “Did you read the paper this morning?”, AD patients tend to respond “Yes, I always read the paper in the morning”, reporting a generic autobiographical memory, without specificity [71].

As per our proposed account, individual variability in the accuracy of the self-model may therefore partly explain why some individuals suffer less severe cognitive changes in the presence of AD [72]. This may suggest therapeutic strategies in Alzheimer’s that might assist retrieval, for example increasing the learning rate of the self-model, to compensate for the lack of episodic memory, or aligning the distribution of chosen actions more closely to the existing self model, for example by sticking to a routine.

## Obsessive-compulsive checking

If confabulating patients confuse imagined actions and decisions for real ones, some individuals with obsessive-compulsive disorder (OCD) show the inverse pattern, doubting whether their actions were truly executed or merely imagined. While findings are mixed with respect to general memory deficits in OCD [73–79], a relatively consistent finding is that obsessive compulsive individuals, especially those who engage in excessive checking, show selectively poor memory of their motor actions [75,78,79, but see also 73], selectively lower levels of confidence in their memory of their motor actions [73,75,77,79], and abnormal sense of agency for the generated actions [80–82]. Perfectly mirroring the typical content of confabulations, obsessive doubts most often involve habitual actions such as turning off the stove or locking the door.

It has been suggested that this specificity to memory for self-actions may be due to the critical function of *reality monitoring* in telling between executed actions and actions that were merely planned or imagined, with some studies reporting poorer ability of obsessive compulsive individuals to distinguish between the two in controlled settings [77,79, but see also 83]. According to this *reality monitoring failure* proposal, obsessive-compulsive individuals learn to mistrust their memory because their imagined actions appear as vivid as their executed ones [75]. This account mirrors predominant theories of confabulation, in which confabulation results from a failure to suppress irrelevant information [61]. Indeed, one study found that sub-clinical checking behaviour was associated with a tendency to falsely remember imagined actions as if they were real, whereas clinically diagnosed obsessive-compulsive individuals tended to commit the inverse error, falsely labelling real actions as merely imagined [75]. This may suggest that obsessive compulsive checking reflects a learned distrust in memory for self-actions.

Recently, an association has been documented between obsessive compulsive disorder and a tendency for dissociative experiences such as depersonalization and maladaptive daydreaming. While this empirical association may reflect various underlying mechanisms [84], it is likely that a memory failure plays a role. Since actions that are taken under a dissociative state may feel automatic and lacking in a “feeling of doing” [85], memory traces for such actions may appear faint and indistinguishable from memory traces for planned or imagined actions [84]. This aligns with the finding of less vivid autobiographical imagery in obsessive-compulsive individuals [86]. Similar to amnesic patients, individuals who regularly dissociate will therefore also depend more on the self-model for reconstructing their past actions.

Indeed, individuals with a tendency for dissociative absorption — a form of dissociation — were less able to recognise words they wrote in a preceding writing session, but performed similarly to control in other memory tests [87,88]. Moreover, acting on “auto-pilot” when in a dissociative state may paradoxically lead to an inflated sense of agency: an individual may recover from a state of dissociation to discover that actions that they only remember planning are now completed. This way, a failure to remember one’s actions may lead to thought-action fusion [89,90], and eventually to the magical thinking characteristic of OCD (“things can happen because I think about them”) [91].

Together, confabulation and obsessive compulsive checking both involve the blurring of a boundary between model-based simulations of likely actions and memory traces of executed actions. As a result, they predominantly feature habits and routines: both as major themes of confabulations, and as a target for obsessive doubt and uncertainty in OCD.

## Concluding remarks

Episodic memory is a record not only of things that happened to us, but also of actions that we took and decisions we made. While the great majority of memory research thus far has focused on memory for facts about the environment, we argue that research into memory for actions and decisions is crucial for developing a full understanding of episodic memory: its involvement in learning about the world and the self, its reliance on schemas and models, and the many ways it can fail in everyday life and in certain pathologies. We propose that episodic memory for one’s actions uniquely relies on access to a self-model, and that, given the role of mental simulations in memory reconstruction and planning, it can lead to reality monitoring failures, making it a particularly vulnerable locus for confabulations on the one hand, and obsessive doubt on the other. Finally, an accurate self-model may cushion against some of the effects of age-related memory loss, hinting at potential therapeutic strategies.

## Box 1: memory for motor actions and routes

While measuring memory for presented stimuli is the rule in the episodic memory literature, some studies do go beyond a passive intake of presented information. Memory for routes, for example, is often measured after participants are allowed to actively explore the environment [e.g., 92–94]. In these studies, however, participants' memory is measured not for the route which they took when exploring the arena (a measure of their memory for their own behaviour), but for their ability to navigate to a target by following the shortest path (an indirect measure of the quality of their mental spatial map). Other experiments that do measure participants' ability to remember their own performed actions do so by instructing them to perform a particular action, later measuring their recall and recognition. Such studies often reveal evidence for an "enactment effect": a memory advantage for performed over observed actions [2–5]. These studies take an important step toward situating episodic memory within a sensorimotor loop. Critically, however, instructing participants to behave in a certain way prevents them from later relying on the self-model to reconstruct their actions: it makes it impossible for them to reason, for example, that "I would not have placed the pen in the cup", because the experimenter may well have instructed me to do so. For this reason, memory for instructed actions is still far from the way to-be-remembered information presents itself in the real world, which is deeply entangled with our voluntary decisions and agentic actions.

## Box 2: Memory for self-actions and retrospective accounts of volition

The inferential process that supports retrospective reconstruction of self-actions may also support the computations of the contingent mental states that accompanied the action in real time: the intention to act and the sense of agency over acting. Indeed, Wegner's Theory of Apparent Mental Causation [37,95] formalised this exact idea. According to this theory, the experience of will can be seen as an inference about causal relations between thoughts and actions. The feeling of having intended an act arises when three cues align: priority (the thought precedes the action), consistency (the thought matches the action), and exclusivity (no alternative cause is evident). In this framework, the experience of will is a form of retrospective sense-making: an inference that arises whenever mental and motor events can be plausibly linked.

Critically, the self-model plays a key role in this story. When introspective evidence is ambiguous or degraded, the self-model provides priors over likely actions, and a sense of having willed the event emerges from the fit between these priors and retrieved outcomes, not from a stored trace of a volitional command. Wegner and Wheatley [95] demonstrated that participants could experience illusory will over actions they did not control when a relevant thought preceded and matched the observed movement. Similarly, covertly altering visual feedback from self-generated cursor or joystick movements leads to a reduction in the experienced sense of agency, and the effect scales with the degree of perturbation [96–98]. These findings can be cast as a real-time analogue of the reconstructive mechanisms that operate when remembering one's own actions.

Recent behavioural evidence from a problem-solving study using the Tower of London task provides a complementary demonstration (Seghezzi et al., in preparation). Participants were more likely to generate false memories of having executed configurations that were valid steps toward the goal than visually similar but route-impossible lures. This selective inflation of endorsements for goal-consistent configurations aligns with a reconstructive account: when mnemonic evidence is sparse, the generative self-model supplies priors over actions that are consistent with the intended goal, biasing retrospective judgements toward plausible moves. In this sense, the result reflects a retrospective reconstruction of agency that is constrained by model-based representations of volition: participants infer not only what they did, but what they must have intended to do. In Wegner's terms, these configurations satisfy consistency and often exclusivity, prompting an apparent sense of having willed or executed the step even when it was never performed.

## Box 3: Self-serving biases in memory for one's decisions

Memory for self-actions is particularly vulnerable to self-serving biases, and specifically to "motivated forgetting". For example, people systematically overestimate how much money they allocated to their game partner in a dictator game, even when incentivised to be accurate [99,100] — an effect that is driven by a selective forgetting of selfish decisions [100]. Similarly, Chew and colleagues [101] had participants complete Raven's Progressive Matrices IQ test, and recall their answers to individual questions several months later. Participants were systematically biased to misremember their responses as more accurate than they were, to report never having seen questions if they did not solve them correctly in the first session, and to recall having answered correctly questions that they had in fact never encountered. In ongoing research (Sawa et al., in preparation), we find that similar self-serving biases in memory for self actions can be observed within a single 15 minute session: participants better remember their correct responses compared to their incorrect ones, and are more confident in their memory of their past guesses if these guesses turned out to be correct. These biases underscore the interdependence of memory for self-actions and the self-model. An overly positive self-model may lead individuals to reconstruct their own past actions as more positive (generous, or accurate) than they in fact were. In turn, selective forgetting of actions that are inconsistent with the self-model may be crucial for maintaining such a positively biased model in the first place.

1. Asperholm, M. *et al.* (2019) What did you do yesterday? A meta-analysis of sex differences in episodic memory. *Psychol. Bull.* 145, 785–821
2. Engelkamp, J. *et al.* (1994) Memory of self-performed tasks: Self-performing during recognition. *Mem. Cognit.* 22, 34–39
3. Kausler, D.H. and Lichty, W. (1988) Memory for Activities: Rehearsal-Independence and Aging. In *Cognitive Development in Adulthood: Progress in Cognitive Development Research* (Howe, M. L. and Brainerd, C. J., eds), pp. 93–131, Springer
4. Roberts, B.R.T. *et al.* (2022) The enactment effect: A systematic review and meta-analysis of behavioral, neuroimaging, and patient studies. *Psychol. Bull.* 148, 397–434
5. Zalla, T. *et al.* (2010) Memory for Self-Performed Actions in Individuals with Asperger Syndrome. *PLOS ONE* 5, e13370
6. Wilkinson, C. and Hyman JR, I.E. (1998) Individual differences related to two types of memory errors: word lists may not generalize to autobiographical memory. *Appl. Cogn. Psychol.* 12, S29–S46
7. Prince, M. (2004) Does Active Learning Work? A Review of the Research. *J. Eng. Educ.* 93, 223–231
8. Theobald, E.J. *et al.* (2020) Active learning narrows achievement gaps for underrepresented students in undergraduate science, technology, engineering, and math. *Proc. Natl. Acad. Sci.* 117, 6476–6483
9. Yannier, N. *et al.* (2021) Active learning: “Hands-on” meets “minds-on.” *Science* 374, 26–30
10. Gershman, S.J. and Daw, N.D. (2017) Reinforcement Learning and Episodic Memory in Humans and Animals: An Integrative Framework. *Annu. Rev. Psychol.* 68, 101–128
11. Gilboa, I. and Schmeidler, D. (2001) *A Theory of Case-Based Decisions*, Cambridge University Press
12. McDougle, S.D. and Hillman, H. (2025) Motor working memory. *Trends Cogn. Sci.* 0
13. Smyth, M.M. and Pendleton, L.R. (1989) Working Memory for Movements. *Q. J. Exp. Psychol. Sect. A* 41, 235–250
14. Hillman, H. *et al.* (2025) Linking motor working memory to explicit and implicit motor learning. *J. Neurophysiol.* 134, 2036–2046
15. Attaallah, B. *et al.* (2024) The role of the human hippocampus in decision-making under uncertainty. *Nat. Hum. Behav.* 8, 1366–1382
16. Biderman, N. *et al.* (2020) What Are Memories For? The Hippocampus Bridges Past Experience with Future Decisions. *Trends Cogn. Sci.* 24, 542–556
17. Ritchie, T.D. *et al.* (2015) A pancultural perspective on the fading affect bias in autobiographical memory. *Memory* 23, 278–290
18. Skowronski, J.J. *et al.* (2014) Chapter Three - The Fading Affect Bias: Its History, Its Implications, and Its Future. In *Advances in Experimental Social Psychology* 49 (Olson, J. M. and Zanna, M. P., eds), pp. 163–218, Academic Press
19. Walker, W.R. *et al.* (2003) Life is Pleasant—and Memory Helps to Keep it that Way! *Rev. Gen. Psychol.* 7, 203–210
20. Breckler, S.J. (1994) Memory for the Experience of Donating Blood: Just How Bad Was It? *Basic Appl. Soc. Psychol.* 15, 467–488
21. Safer, M.A. *et al.* (2002) Distortion in Memory for Emotions: The Contributions of Personality and Post-Event Knowledge. *Pers. Soc. Psychol. Bull.* 28, 1495–1507
22. Levine, L.J. (1997) Reconstructing memory for emotions. *J. Exp. Psychol. Gen.* 126, 165–177
23. Fredrickson, B.L. and Kahneman, D. (1993) Duration neglect in retrospective evaluations of affective episodes. *J. Pers. Soc. Psychol.* 65, 45–55

24. Cushman, F. (2020) Rationalization is rational. *Behav. Brain Sci.* 43, e28
25. Bartlett, S.F.C. (1995) *Remembering: A Study in Experimental and Social Psychology*, Cambridge University Press
26. Levin, M. (2024) Self-improvising Memory: a perspective on memories as agential, dynamically-reinterpreting cognitive glueOSF
27. Spens, E. and Burgess, N. (2024) A generative model of memory construction and consolidation. *Nat. Hum. Behav.* 8, 526–543
28. Hemmer, P. and Steyvers, M. (2009) A Bayesian Account of Reconstructive Memory. *Top. Cogn. Sci.* 1, 189–202
29. De Brigaard, Felipe Remembering as inverse causal inference. *Philos. Psychol.* at <https://www.tandfonline.com/doi/abs/10.1080/09515089.2026.2650495>
30. Chi, M.T.H. and Ceci, S.J. (1987) Content Knowledge: Its Role, Representation, and Restructuring in Memory Development. In *Advances in Child Development and Behavior* 20 (Reese, H. W., ed), pp. 91–142, JAI
31. Schneider, W. *et al.* (1993) Chess Expertise and Memory for Chess Positions in Children and Adults. *J. Exp. Child Psychol.* 56, 328–349
32. Rogers, T.B. *et al.* (1977) Self-Reference and the Encoding of Personal Information. *J. Pers. Soc. Psychol.* 35
33. Symons, C.S. and Johnson, B.T. (1997) The self-reference effect in memory: A meta-analysis. *Psychol. Bull.* 121, 371–394
34. Gopnik, A. (1993) How we know our minds: The illusion of first-person knowledge of intentionality. *Behav. Brain Sci.* 16, 1–14
35. Johansson, P. *et al.* (2006) How something can be said about telling more than we can know: On choice blindness and introspection. *Conscious. Cogn.* 15, 673–692
36. Nisbett, R.E. and Wilson, T.D. (1977) Telling more than we can know: Verbal reports on mental processes. *Psychol. Rev.* 84, 231–259
37. Wegner, D.M. (2003) The Mind's Self-Portrait. *Ann. N. Y. Acad. Sci.* 1001, 212–225
38. Craik, K.J.W. (1967) *The nature of explanation*, 445, CUP Archive
39. Lake, B.M. *et al.* (2015) Human-level concept learning through probabilistic program induction. *Science* 350, 1332–1338
40. Flavell, J.H. (1979) Metacognition and cognitive monitoring: A new area of cognitive–developmental inquiry. *Am. Psychol.* 34, 906–911
41. Marr, D. (2010) *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*, MIT Press
42. Gilboa, A. and Marlatte, H. (2017) Neurobiology of Schemas and Schema-Mediated Memory. *Trends Cogn. Sci.* 21, 618–631
43. McClelland, J.L. *et al.* (1995) Why there are complementary learning systems in the hippocampus and neocortex: Insights from the successes and failures of connectionist models of learning and memory. *Psychol. Rev.* 102, 419–457
44. Winocur, G. *et al.* (2010) Memory formation and long-term retention in humans and animals: Convergence towards a transformation account of hippocampal–neocortical interactions. *Neuropsychologia* 48, 2339–2356
45. Denny, B.T. *et al.* (2012) A Meta-analysis of Functional Neuroimaging Studies of Self- and Other Judgments Reveals a Spatial Gradient for Mentalizing in Medial Prefrontal Cortex. *J. Cogn. Neurosci.* 24, 1742–1752
46. Iravani, B. *et al.* (2024) Intracranial Recordings of the Human Orbitofrontal Cortical Activity during Self-Referential Episodic and Valenced Self-Judgments. *J. Neurosci.* 44
47. Northoff, G. *et al.* (2011) Brain imaging of the self – Conceptual, anatomical and methodological issues. *Conscious. Cogn.* 20, 52–63
48. Stendardi, D. *et al.* (2023) Who am I really? The ephemerality of the self-schema following vmPFC damage. *Neuropsychologia* 188, 108651

49. Yin, S. *et al.* (2021) Ventromedial Prefrontal Cortex Drives the Prioritization of Self-Associated Stimuli in Working Memory. *J. Neurosci.* 41, 2012–2023
50. Schurz, M. *et al.* (2014) Fractionating theory of mind: A meta-analysis of functional brain imaging studies. *Neurosci. Biobehav. Rev.* 42, 9–34
51. Ma, F. *et al.* (2025) Prediction, inference, and generalization in orbitofrontal cortex. *Curr. Biol.* 35, R266–R272
52. Moneta, N. *et al.* (2024) Representational spaces in orbitofrontal and ventromedial prefrontal cortex: task states, values, and beyond. *Trends Neurosci.* 47, 1055–1069
53. Wang, F. *et al.* (2020) Interactions between human orbitofrontal cortex and hippocampus support model-based inference. *PLOS Biol.* 18, e3000578
54. Nilsson, L.-G. *et al.* (2000) Activity in motor areas while remembering action events. *NeuroReport* 11, 2199
55. Marr, D. (1971) Simple memory: a theory for archicortex. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 262, 23–81
56. Norman, K.A. and O'Reilly, R.C. (2003) Modeling hippocampal and neocortical contributions to recognition memory: A complementary-learning-systems approach. *Psychol. Rev.* 110, 611–646
57. Borsutzky, S. *et al.* (2008) Confabulations in alcoholic Korsakoff patients. *Neuropsychologia* 46, 3133–3143
58. Bateman, J.R. *et al.* (2024) Network Localization of Spontaneous Confabulation. *J. Neuropsychiatry Clin. Neurosci.* 36, 45–52
59. Burgess, P.W. and McNeil, J.E. (1999) Content-Specific Confabulation. *Cortex* 35, 163–182
60. Dalla Barba, G. *et al.* (2017) A longitudinal study of confabulation. *Cortex* 87, 44–51
61. Schnider, A. (2003) Spontaneous confabulation and the adaptation of thought to ongoing reality. *Nat. Rev. Neurosci.* 4, 662–671
62. Dalla Barba, G. and Boissé, M.-F. (2010) Temporal consciousness and confabulation: Is the medial temporal lobe "temporal"? *Cognit. Neuropsychiatry* 15, 95–117
63. Gilboa, A. *et al.* (2002) The cognitive neuroscience of confabulation: A review and a model. *Handb. Mem. Disord.* 2, 315–342
64. Kessels, R.P.C. and Kopelman, M.D. (2012) Context Memory in Korsakoff's Syndrome. *Neuropsychol. Rev.* 22, 117–131
65. Huppert, F.A. and Piercy, M. (1976) Recognition Memory in Amnesic Patients: Effect of Temporal Context and Familiarity of Material. *Cortex* 12, 3–20
66. Mazor, M. *et al.* (2024) Pretending not to know reveals a powerful capacity for self-simulation. DOI: 10.31234/osf.io/pgxrz
67. Kossale, Y. *et al.* (2022) Mode Collapse in Generative Adversarial Networks: An Overview. in *2022 8th International Conference on Optimization and Applications (ICOA)*, pp. 1–6
68. Schoemaker, D. *et al.* (2014) Recollection and Familiarity in Aging Individuals with Mild Cognitive Impairment and Alzheimer's Disease: A Literature Review. *Neuropsychol. Rev.* 24, 313–331
69. Giffard, B. *et al.* (2001) The nature of semantic memory deficits in Alzheimer's disease: New insights from hyperpriming effects. *Brain* 124, 1522–1532
70. Ober, B.A. *et al.* (1995) Assessment of associative relations in alzheimer's disease: Evidence for preservation of semantic memory. *Aging Neuropsychol. Cogn.* 2, 254–267
71. Philippi, N. *et al.* (2015) Different Temporal Patterns of Specific and General Autobiographical Memories across the Lifespan in Alzheimer's Disease. *Behav. Neurol.* 2015, 963460
72. M. Tucker, A. and Stern, Y. (2011) Cognitive Reserve in Aging. *Curr. Alzheimer Res.* 8, 354–360

73. Cougle, J.R. *et al.* (2008) “Perhaps you only imagined doing it”: reality-monitoring in obsessive-compulsive checkers using semi-idiographic stimuli. *J. Behav. Ther. Exp. Psychiatry* 39, 305–320
74. Dar, R. *et al.* (2022) Are people with obsessive-compulsive disorder under-confident in their memory and perception? A review and meta-analysis. *Psychol. Med.* 52, 2404–2412
75. Ecker, W. and Engelkamp, J. (1995) Memory for Actions in Obsessive-Compulsive Disorder. *Behav. Cogn. Psychother.* 23, 349–371
76. Exner, C. *et al.* (2009) Metacognition and episodic memory in obsessive-compulsive disorder. *J. Anxiety Disord.* 23, 624–631
77. McNally, R.J. and Kohlbeck, P.A. (1993) Reality monitoring in obsessive-compulsive disorder. *Behav. Res. Ther.* 31, 249–253
78. Sher, K.J. *et al.* (1989) Memory deficits in compulsive checkers: Replication and extension in a clinical sample. *Behav. Res. Ther.* 27, 65–69
79. Zermatten, A. *et al.* (2006) Reality monitoring and motor memory in checking-prone individuals. *J. Anxiety Disord.* 20, 580–596
80. De Haan, S. *et al.* (2015) Being free by losing control: what obsessive-compulsive disorder can tell us about free will
81. Oren, E. *et al.* (2019) Intentional binding and obsessive-compulsive tendencies: A dissociation between indirect and direct measures of the sense of agency. *J. Obsessive-Compuls. Relat. Disord.* 20, 59–65
82. Gentsch, A. *et al.* (2012) Dysfunctional Forward Model Mechanisms and Aberrant Sense of Agency in Obsessive-Compulsive Disorder. *Biol. Psychiatry* 71, 652–659
83. Merckelbach, H. and Wessel, I. (2000) Memory for Actions and Dissociation in Obsessive-Compulsive Disorder. *J. Nerv. Ment. Dis.* 188, 846
84. Soffer-Dudek, N. (2023) Obsessive-compulsive symptoms and dissociative experiences: Suggested underlying mechanisms and implications for science and practice. *Front. Psychol.* 14
85. Belayachi, S. and Van der Linden, M. (2010) Feeling of doing in obsessive–compulsive checking. *Conscious. Cogn.* 19, 534–546
86. Zermatten, A. *et al.* (2008) Phenomenal characteristics of autobiographical memories and imagined events in sub-clinical obsessive-compulsive checkers. *Appl. Cogn. Psychol.* 22, 113–125
87. Bregman-Hai, N. *et al.* (2020) Who wrote that? Automaticity and reduced sense of agency in individuals prone to dissociative absorption. *Conscious. Cogn.* 78, 102861
88. Chiu, C.-D. *et al.* (2016) Misattributing the Source of Self-Generated Representations Related to Dissociative and Psychotic Symptoms. *Front. Psychol.* 7
89. Rees, C.S. *et al.* (2010) The Relationship Between Magical Thinking, Thought-Action Fusion and Obsessive-Compulsive Symptoms. *Int. J. Cogn. Ther.* 3, 304–311
90. Shafran, R. *et al.* (1996) Thought-action fusion in obsessive compulsive disorder. *J. Anxiety Disord.* 10, 379–391
91. Einstein, D.A. and Menzies, R.G. (2004) The presence of magical thinking in obsessive compulsive disorder. *Behav. Res. Ther.* 42, 539–549
92. Sneider, J.T. *et al.* (2011) A Preliminary Study of Sex Differences in Brain Activation during a Spatial Navigation Task in Healthy Adults. *Percept. Mot. Skills* 113, 461–480
93. Lövdén, M. *et al.* (2012) Spatial navigation training protects the hippocampus against age-related changes during early and late adulthood. *Neurobiol. Aging* 33, 620.e9-620.e22
94. Lövdén, M. *et al.* (2005) Environmental topography and postural control demands shape aging-associated decrements in spatial navigation performance. *Psychol. Aging* 20, 683–694
95. Wegner, D.M. and Wheatley, T. (1999) Apparent mental causation: Sources of the experience of will. *Am. Psychol.* 54, 480–492

96. Farrer, C. *et al.* (2003) Modulating the experience of agency: a positron emission tomography study. *NeuroImage* 18, 324–333
97. Farrer, C. *et al.* (2008) The Angular Gyrus Computes Action Awareness Representations. *Cereb. Cortex* 18, 254–261
98. Farrer, C. and Frith, C.D. (2002) Experiencing Oneself vs Another Person as Being the Cause of an Action: The Neural Correlates of the Experience of Agency. *NeuroImage* 15, 596–603
99. Carlson, R.W. *et al.* (2020) Motivated misremembering of selfish decisions. *Nat. Commun.* 11, 2100
100. Saucet, C. and Villeval, M.C. (2019) Motivated memory in dictator games. *Games Econ. Behav.* 117, 250–275
101. Chew, S.H. *et al.* (2020) Motivated False Memory. *J. Polit. Econ.* 128, 3913–3939

## Highlights

- We name three key functions that memory for self-actions uniquely plays in human cognition:
    - It scaffolds memory for other aspects of the environment ("*I remember cycling through the busy town center, so the river path must have been flooded*").
    - It allows learning in settings where feedback is delayed ("*I remember studying only until dinner before the day of the exam, and I got an A*").
    - It makes it possible to learn about one's own subjective preferences from observable actions ("*I remember staring at the ceiling during the movie, so I must have found it boring*").

- Contemporary accounts of episodic memory assume a key role for cortically stored schemas and models. Building on this work, we suggest a framework in which memory for our actions critically relies on a self-model: a model which specifies how we believe we would behave in different settings. The question "*where did I park the car?*" quickly becomes "*where would I park the car?*" in the absence of a clear memory trace. For this reason, a complete account of episodic memory should feature not only abstract knowledge about the external world, but also metacognitive knowledge about one's own cognition.

- We extend our framework to consider failure of memory for self actions in pathological confabulation (for example, in Korsakoff syndrome), over-reliance on generic habits and schemas (for example, in Alzheimer's disease), and a selective under-confidence in memory for self-actions (in Obsessive Compulsive Disorder).