

The Write Kind of Memory: A Functionalist Critique of the Extended Mind Hypothesis

TESSA POPPLE¹

UNIVERSITY OF QUEENSLAND

Abstract

This paper offers a functionalist critique of the Extended Mind Hypothesis as advanced by Andy Clark and David Chalmers in response to the question: “where is my mind?” While accepting the core functionalist premise that cognitive status is determined by causal role rather than location, it argues that the analogical Otto/Inga case fails to establish genuine functional parity. Thus, the Extended Mind Hypothesis fails. The argument draws on contemporary accounts of memory as a reconstructive and epistemically regulative process, supplemented by Eric Schwitzgebel's work on the unreliability of introspection, to argue that biological memory possesses functionally significant features absent from Otto's notebook. The paper concludes that despite surface similarities in the Otto/Inga analogy, a disanalogy can be drawn when the functionalism Clark and Chalmers depend on is truly taken seriously.

¹ Tessa Popple is a third-year undergraduate student who is undertaking an extended major in philosophy at the University of Queensland. Her interests include the philosophy of mind and cognitive science, epistemology, and philosophy of logic.

§ 1. From Extension to Objection

Contemporary philosophy of mind is increasingly shaped by the suspicion that the boundaries of the mental are far less clear than once assumed. Traditional images of an inner, self-contained mind have given way to models that emphasise embodiment, environmental coupling, and the pervasive role of external resources in cognition. Against this backdrop, Andy Clark and David Chalmers' (C&C) extended mind hypothesis (EMH) proposes that cognitive processes are not confined behind the skin-and-skull barrier but can extend into the world through our routine use of tools.² Their central case is the now-canonical analogy between Otto and Inga: Inga recalls information from biological memory, while Otto, who suffers from Alzheimer's disease, consults a notebook in which he has recorded the same information. When the notebook is reliably available, automatically trusted, and seamlessly integrated into Otto's activity, C&C argue that it plays the same functional role as memory, and should therefore be counted as part of his cognitive system.³ The force of this claim depends on a commitment to functionalism: that what makes a process cognitive is the role it plays within a system, rather than its physical realisation or location.⁴ If Otto's notebook performs the same function as Inga's memory, then denying it cognitive status would amount to privileging the inside of the head in a theoretically unmotivated way.⁵ The EMH thus presents itself not as a radical departure from familiar assumptions, but as a natural consequence of taking functionalism seriously. In this paper, I argue that the Otto/Inga analogy fails to establish sufficient functional parity to support the EMH precisely because functionalism is taken seriously. The problem lies not in the appeal to external resources as such, but in the account of memory on which the analogy depends. C&C treat memory as if it were a passive store of information, whose role is exhausted by the reliable retrieval of stored content. Biological memory extends beyond storage, functioning as a reconstructive capacity that operates under conditions of uncertainty.⁶ Its characteristic fallibility — a tendency to distort, misrepresent, and revise — is not an incidental defect, but a functionally essential feature of how it supports reasoning, learning, and action.⁷ The appeal to subjectivity in this context is functional rather than phenomenological. It concerns the way memory regulates an agent's epistemic stance by presenting information with varying degrees of confidence, salience, and authority, not the qualitative character of remembering itself. Drawing on empirical work in the neuropsychology of memory and Eric Schwitzgebel's account of introspection, I argue that this reconstructive and epistemically regulative character is constitutive of biological

² Clark, Andy, and David Chalmers (1998), 'The Extended Mind', *Analysis* 58: 7.

³ Clark and Chalmers, *The Extended Mind*, 2–16.

⁴ Clark and Chalmers, *The Extended Mind*, 8.

⁵ Clark and Chalmers, *The Extended Mind*, 8.

⁶ Schacter, Daniel L (2012) 'Adaptive Constructive Processes and the Future of Memory', *The American Psychologist* 67: 603–13.

⁷ Bartlett, Frederic C (1995) 'Perceiving, Recognising, Remembering', in *Remembering*: 186–96: Cambridge University Press.

remembering in a way that is not captured by Otto's notebook. I conclude that even if the mind extends, Otto and Inga do not provide sufficient evidence of it.

§ 2. Mind the Gap: Cognition Beyond the Skull

C&C's EMH is motivated by a substantive commitment to functionalism: the view that what makes a state mental is not what it is made of, nor where it is located, but the role it plays within a system of inputs, internal processes, and behavioural outputs. On this approach, mental states are individuated by their causal roles — how they are formed, how they interact with other states, and how they guide action — rather than by their physical realisation. While functionalism can encompass a range of positions, C&C rely on the broadly shared commitment that cognitive states are individuated by their functional roles rather than by the material in which they are realised. This allows for the possibility that the same mental state could, in principle, be realised in different forms, provided the relevant functional organisation is preserved. The EMH extends this commitment by rejecting the assumption that such functional organisation must be realised entirely within the biological boundaries of the individual. This motivates what is known as the parity principle for C&C: if a process outside the head plays the same explanatory role as one inside it, then to deny it cognitive status on the basis of location is arbitrary and theoretically unmotivated.⁸ The principle is designed to block what they take to be an unjustified 'skin-and-skull bias' in our intuitions about the mind.⁹ If we are willing to count a process as cognitive when it is realised neurally, consistency demands that we extend the same status to non-neural processes that perform the same function. This position reflects a broader reorientation within cognitive science, in which agents are understood not as self-contained information processors but as systems dynamically coupled to their environments. Cognition, on this view, is not exhausted by what happens internally, but is distributed across brain, body, and world. Human problem-solving routinely involves the active manipulation of external resources such as using a calculator to perform complex arithmetic, rearranging symbols on paper to solve a problem, or relying on digital devices to store and organise information. These practices are not merely convenient aids but are deeply integrated into our cognitive routines, shaping how tasks are structured and how solutions are reached. The EMH thus promises a convincing continuity between inner processes and our pervasive use of tools, presenting cognitive extension as a natural consequence of taking functionalism and our everyday practices seriously.

Traditionally, philosophy of mind has regarded cognitive processes as occurring wholly within the individual, whether understood in terms of brain states,

⁸ Clark and Chalmers, *The Extended Mind*, 8–9.

⁹ Clark and Chalmers, *The Extended Mind*, 14.

representations, or subjective experience.¹⁰ This is the internal view of cognition which externalism rejects, arguing that the contents of our thoughts can depend in part on factors outside the individual. One such challenge to internalism is found in Hilary Putnam's *The meaning of 'meaning'*.¹¹ Through his Twin Earth thought experiment, Putnam argues that the meanings of our terms are not determined solely by our internal psychological states, since two individuals could be psychologically identical while nevertheless referring to different substances when they use the word "water".¹² On this view, features of the external environment partly determine mental content, leading Putnam to conclude that "meanings just ain't in the head."¹³

Whereas Putnam emphasises dependence on the physical environment, Tyler Burge develops this in *Individualism and the Mental* by arguing that what an individual thinks is partly determined by the linguistic community in which they participate.¹⁴ He presents a thought experiment that aims to show that two individuals could be physically and psychologically identical, yet differ in what they think due solely to differences in how terms are used in their respective linguistic communities.¹⁵ On this view, referred to as *passive* externalism, the environment plays a constitutive role in fixing the contents of mental states, but does so without locating those states themselves: beliefs, thoughts, and intentions remain states of the subject, even if what they are about depends on external factors.¹⁶

C&C both extend and depart from this framework by hypothesising *active* externalism. This goes further than the argument that the environment helps determine what we think, instead asserting that it can be directly involved in the processes by which we think.¹⁷ Whereas passive externalism leaves cognition located within the individual while allowing its content to be environmentally shaped, C&C challenge the assumption that cognitive processes must be confined within the boundaries of skin and skull: the claim is not that the world helps determine what we think, but that it can literally participate in some of the thinking.¹⁸ The EMH therefore asks us to accept not just that cognition is environmentally dependent, but that the environment itself forms part of cognition.

¹⁰ See: Descartes, René (1998) *Discourse on Method and Meditations on First Philosophy*, 4th edition, Donald Cress, trans. Hackett; Fodor, Jerry A (1975) *The Language of Thought. Language and Thought Series*. Crowell, Adams, Fred, and Kenneth Aizawa (2001) 'Why the Mind Is Still in the Head', in Philip Robbins and Murat Aydede, eds., *The Cambridge Handbook of Situated Cognition*: 78–95. Cambridge University Press.

¹¹ Putnam, Hilary (2011) 'The Meaning of 'Meaning'', in Hilary Putnam *Mind, Language and Reality*: 215–71. Cambridge University Press.

¹² Putnam, *The meaning of 'meaning'*, 223–26.

¹³ Putnam, *The meaning of 'meaning'*, 227.

¹⁴ Burge, Tyler (1979) 'Individualism and the Mental', *Midwest Studies in Philosophy* 4: 73–121.

¹⁵ Burge, *Individualism and the Mental*, 94.

¹⁶ Burge, *Individualism and the Mental*, 79.

¹⁷ Clark and Chalmers, *The Extended Mind*, 9.

¹⁸ Clark and Chalmers, *The Extended Mind*, 9.

§ 3. An Analogy from Memory to Memo

This is the backdrop against which C&C introduce their central thought experiment, designed to make the extended mind hypothesis intuitive and sensitive to our daily practices. Having argued that cognition often involves the active use of environmental resources — such as manipulating objects, writing, or rearranging external symbols — they turn from general cases of problem-solving to more specific cases of belief and memory. Their aim is to show that external resources can play the same functional role as internal cognitive states and therefore qualify as constituents of cognition. Where sufficient functional parity exists, differences in physical location do not justify treating one process as cognitive and the other as merely supportive. To this end, they contrast two cases analogically.¹⁹ Inga is an ordinary agent with standard cognitive capacities. Upon hearing that an exhibition is on at the Museum of Modern Art, she forms the intention to visit and recalls from her biological memory that the museum is on 53rd Street. Prior to recall, the information exists as a standing belief stored in memory, available when needed to guide action. When she retrieves it, she acts on it without hesitation and walks to the appropriate location. Otto, by contrast, suffers from Alzheimer's disease and cannot reliably store or retrieve information biologically. To compensate, he relies on an external artifact: a notebook in which he records information as he encounters it. When Otto learns about the exhibition, he writes down the address. Later, when he decides to visit, he consults the notebook, finds the relevant entry, and proceeds accordingly. The notebook is used by Otto as a constant and trusted resource — “the information is reliably there when needed, available to consciousness and available to guide action”.²⁰ C&C argue that once these conditions are in place, there is no principled difference between the two cases. In both cases, the information is reliably available, easily accessible, and directly guides action in response to the agent's goals. To deny that Otto believes the museum is on 53rd Street, while granting that Inga does, would be to privilege the location of the information as inside the head over its functional role. This is precisely what their parity principle is designed to reject.

To avoid ‘cognitive bloat’, a concern expressed by critics of the EMH in which the boundaries of the mind become too permissive, C&C identify a set of conditions under which an external resource qualifies as part of a cognitive process: it must be reliably available, easily accessible, automatically endorsed, and integrated into the agent's ongoing activity.^{21 22} Otto's notebook, on this view, plays the same functional role for him as biological memory plays for Inga. The analogy is therefore not between a person and a person-with-a-tool, but between two cognitive systems that differ only in the location of their informational resources. If functionalism is accepted, C&C conclude, then parity is achieved. The notebook therefore constitutes Otto's memory itself, extending cognition beyond the skin-and-skull boundary.

¹⁹ Clark and Chalmers, *The Extended Mind*, 13–14.

²⁰ Clark and Chalmers, *The Extended Mind*, 13–14.

²¹ Clark and Chalmers, *The Extended Mind*, 11.

²² Sprevak, Mark (2019) ‘Extended Cognition’, *The Routledge Encyclopedia of Philosophy Online*: 1–15.

C&C anticipate a range of objections to the claim of parity in the analogy, many of which arise from the intuitive sense that Inga's case involves a more direct or genuine form of cognition than Otto's.²³ These objections take several forms. One concerns access: Inga appears to enjoy a privileged, immediate relation to her memory, whereas Otto must consult an external object. Another concerns mode of access: Otto's interaction with the notebook is perceptual as he reads what is written, whereas Inga's access to memory is often characterised as introspective or non-perceptual. A further objection targets the nature of belief itself, suggesting that only occurrent mental states — those currently present to consciousness — qualify as genuine beliefs. In this case, Otto's reliance on a stored notebook entry would count only as dispositional or derivative. Finally, there are concerns about reliability and stability, since Otto's notebook might be lost, damaged, or unavailable in ways that biological memory typically is not.²⁴ Despite granting the intuition as to which these objections arise, C&C argue that they ultimately fail to identify a principled difference between the cases. Differences in access — perceptual versus introspective — are dismissed as variations in the mode of information retrieval rather than differences in its functional role. Likewise, concerns about reliability are treated as matters of degree rather than kind: biological memory is itself fallible, subject to interference, decay, and disruption, and thus cannot claim a privileged status on this basis alone. Even the appeal to occurrent belief is rejected as too restrictive, since much of what we ordinarily count as belief consists in standing states that are not continuously present to consciousness but are nonetheless available to guide action when required. In each case, C&C's strategy is to show that the apparent asymmetries between Otto and Inga are superficial — differences in medium, phenomenology, or implementation — rather than differences in the underlying cognitive role played by the relevant information.

Formally, the argument is both conditional and analogical. It proceeds from a functionalist premise: namely, that cognitive status is determined by the role a process plays within a system, and that it applies to an analogy in which an internal and an external process appear to realise that role equally well. If functionalism is true, and if Otto's notebook is functionally equivalent to Inga's biological memory, then the information stored in the notebook must count as a belief or memory state. One influential line of resistance, developed by Fred Adams and Kenneth Aizawa, rejects the sufficiency of functional equivalence altogether.²⁵ They argue that C&C's account lacks a principled criterion for distinguishing genuinely cognitive processes from merely causal accompaniments and propose instead that cognition is marked by features such as the presence of non-derived representations and specific kinds of neural processing.²⁶ On this view, external artifacts like notebooks may play an important role in facilitating cognition but do not themselves constitute cognitive processes, since their representational content is derivative on human interpretation.

²³ Clark and Chalmers, *The Extended Mind*, 14–16.

²⁴ Clark and Chalmers, *The Extended Mind*, 14–16.

²⁵ Adams and Aizawa, *Why the Mind Is Still in the Head*, 78–79.

²⁶ Adams and Aizawa, *Why the Mind Is Still in the Head*, 80.

This objection thus targets the functionalist premise directly, seeking to reinstate a boundary between the cognitive and the non-cognitive that is not captured by parity of role alone.

§ 4. Reconstructing Memory, Deconstructing Parity

We need not reject functionalism in order to resist C&C's conclusion. Indeed, C&C themselves acknowledge their argument leaves open a different strategy of objection: namely, to accept functionalism but deny relevant functional equivalence has been established. As they explicitly concede, to "provide substantial resistance, an opponent has to show that Otto's and Inga's cases differ in some important and relevant aspect."^[1]²⁷ This shifts the burden of the debate. The question is no longer whether cognition must be internal, nor whether functionalism is correct, but whether the Otto/Inga analogy genuinely satisfies the conditions of parity. Parity stands only if no deeper functional disanalogy can be shown.

Otto's and Inga's cases do differ in an important and relevant way once a more nuanced account of memory is adopted. C&C's argument succeeds only if memory is understood as a passive store of information whose function is exhausted by the reliable retrieval of stored content. Biological memory is instead a perspective-bound, reconstructive capacity that retrieves information under conditions of uncertainty, salience, and confidence.²⁸ Memory is fallible *as memory*, distorting and misrepresenting information while often presenting itself as authoritative.²⁹ As Loftus and Pickrell observe, "the memory error occurs because details of experienced events or imagined events are integrated with inferences and other elaborations," a consequence of the ordinary mechanisms through which memories are formed and retrieved.³⁰ These failures are not occasional defects, but internal to the way remembering functions for a subject who stands in a first-person relation to their past. Remembering functions through the continual updating, reorganisation, and evaluation of representations in light of present circumstances. This functional organisation enables flexible reasoning and adaptation while also generating the possibility of distortion. Error therefore arises from the ordinary operation of remembering itself. Inga can be mistaken in recalling an address, certain when she should doubt, or unsure when she is in fact right. This is central to what memory does. Otto's notebook, by contrast, is static and epistemically transparent. It does not reinterpret its contents, reorder them in response to contextual cues, or present them with varying degrees of confidence. It is just as "sure" of what Otto ate for lunch five years ago as it is of the museum's address — a mistaken notebook entry is simply an incorrect record.

²⁷ Clark and Chalmers, *The Extended Mind*, 14.

²⁸ Schacter, *Adaptive Constructive Processes and the Future of Memory*.

²⁹ Loftus, Elizabeth F, and Jacqueline E Pickrell (1995) 'The Formation of False Memories', *Psychiatric Annals* 25: 720–25.

³⁰ Loftus and Pickrell, *The Formation of False Memories*, 724.

Consider a conversation in which Inga's friend reminds her of an event from years ago. Her recollection may subsequently change as the reconstructive mechanisms of memory integrate new contextual information into the remembered event. The resulting revision is produced by memory's ordinary operation rather than by damage or malfunction. By contrast, if Otto's notebook now records that the guest arrived late, this is because the entry has been altered or because Otto originally recorded it differently. In this respect, Otto may even possess a more stable informational resource than Inga, since the notebook preserves fixed content without the reconstructive distortions characteristic of memory.

Biological memory is a norm-governed, perspectival capacity whose characteristic fallibility is internal to its function. Its role includes regulating an agent's epistemic stance by selecting, weighting, and reconstructing representations in light of context, alongside the storage and retrieval of information. Those same functional processes inevitably introduce the possibility of distortion, making fallibility constitutive of memory rather than an accidental defect. The notebook lacks this core feature. This objection does not rest on the claim that Otto's notebook lacks reliability, availability, or integration, nor that it differs merely in medium or location. The relevant distinction concerns the functional organisation of remembering rather than the phenomenology of memory; the appeal to subjectivity here is functional. Biological memory presents information, while also regulating epistemic stance by producing representations graded in confidence, shaped by context, and open to revision. Retrieval is therefore inseparable from the evaluation and reconstruction of remembered content. The same processes that make a memory more or less salient, tentative, or authoritative also integrate new information and can alter the remembered representation itself. By contrast, consulting a notebook retrieves a comparatively stable record whose epistemic significance is determined through subsequent interpretation. Any change to the stored content requires a distinct act of revision rather than emerging through the ordinary process of retrieval itself. Otto and Inga therefore differ in an important and relevant way.

Eric Schwitzgebel's work on the unreliability of introspection provides further support for this claim by challenging a deeply entrenched assumption in the philosophy of mind: that we enjoy a privileged and largely authoritative access to our own mental states. Traditionally, introspection has been treated as a kind of epistemic bedrock, a domain in which error is minimal and self-knowledge secure.³¹ Schwitzgebel objects to this assumption, arguing that even under favourable conditions of careful reflection, introspection is systematically and pervasively unreliable. He contends that we are often "ignorant and prone to error" about our own ongoing conscious experience, not only in marginal or difficult cases, but even with respect to its most basic and pervasive features.³² For example, he argues that most people initially judge themselves to possess a broad field of stable, richly

³¹ See Descartes, *Discourse on method ; and, Meditations on first philosophy.*; Shoemaker, Sydney (1963) *Self-Knowledge and Self-Identity* Cornell University Press.

³² Schwitzgebel, Eric (2008) 'The Unreliability of Naive Introspection', *The Philosophical Review* 117: 245–73.

detailed visual experience. However, when they undertake simple introspective exercises—such as maintaining fixation while attending to peripheral vision—they frequently revise this judgment, recognising that the apparent richness of visual experience was substantially overstated.³³ For Schwitzgebel, this demonstrates that even careful reflection can expose systematic errors in our understanding of our own conscious experience. This unreliability is not limited to judgments about the causes of our mental states, or about dispositional attitudes such as character traits or motivations, where error might be expected. Schwitzgebel's examples are deliberately mundane and wide-ranging. We struggle to characterise the nature of our emotional experience — whether emotions have a stable phenomenological core, how they are spatially or bodily located, or even whether we are undergoing them at all. Similarly, we misjudge the structure of our visual experience, overestimating the clarity and detail of the visual field and failing to appreciate the limited and shifting nature of perceptual attention.³⁴ These mistakes are common and persistent features of ordinary cognition. Even when we reflect sincerely and attentively, our judgments about our own experience are unstable, revisable, and often incorrect. Schwitzgebel concludes that mental states are not self-presenting in the way many philosophical accounts assume.³⁵ Our access to them is mediated, interpretive, and subject to the same kinds of error, uncertainty, and revision that characterise our engagement with the external world. To have a belief or a memory is thus not to stand in a relation of transparent self-knowledge, but to occupy a position that admits of misjudgement about one's own epistemic state. This supports the present argument by reinforcing the idea that fallibility is not a structural feature of cognition: the same processes that enable access to mental states also necessitate the possibility of error in accessing them. The question, then, is whether features of this picture of memory can be attributed to Otto's notebook.

As such, C&C may respond that this line of argument fails to establish a principled disanalogy. That biological memory exhibits a distinctive form of fallibility does not, on their view, mark a relevant difference, since Otto's notebook is not perfectly reliable either: it can be lost, damaged, misread, or even tampered with. In this respect, both systems are susceptible to error, and the source of that error — whether internal or external — does not by itself determine cognitive status. What matters, they argue, is not perfect reliability, nor the presence of a distinctive phenomenology, but whether a resource is sufficiently available, accessible, and integrated into the agent's cognitive routines to play the appropriate action-guiding role.³⁶ On this view, the relevant standard is one of functional integration rather than intrinsic character. Biological memory is itself imperfect, subject to interference, decay, and distortion, yet it nonetheless counts as memory because it is reliably available when needed and

³³ Schwitzgebel, *The Unreliability of Naïve Introspection*, 255–56.

³⁴ Schwitzgebel, *The Unreliability of Naïve Introspection*, 265–67.

³⁵ Schwitzgebel, *The Unreliability of Naïve Introspection*, 268.

³⁶ Clark, Andy (2008) *Supersizing the Mind : Embodiment, Action, and Cognitive Extension*. *Philosophy of Mind*. Oxford University Press.

is seamlessly incorporated into ongoing cognition. If Otto's notebook is similarly dependable, then it satisfies the same functional criteria. That its failures arise through external damages rather than through reconstructive distortion is treated as a difference in implementation, rather than a difference in role. From this perspective, differences in implementation or location do not by themselves justify denying parity, since doing so would risk reinstating the very skin-and-skull boundary that the EMH challenges.

This reply mislocates the source of the distinction. The contrast is not between systems that can err and systems that cannot, but between systems whose *normal operation* includes epistemic risk and those whose errors arise only from malfunction. Fallibility due to loss, corruption, or damage is merely contingent: it reflects what can happen to any physical object. Fallibility due to reconstruction, by contrast, is constitutive of memory. Remembering reconstructs representations of past events through context-sensitive processes that operate from the subject's epistemic perspective. In memory, error is produced by the same mechanisms that make recall possible at all: in a system involving a notebook, error arises through failures in interpretation, retrieval, or integration, rather than from the operation of the storage medium itself.³⁷ This point survives the reply that error may occur at the level of the coupled system. Otto may misread, misinterpret, or misapply what is written just as he may misjudge his own memories. But in such cases, the source of the error lies in the cognitive processing brought to bear on the stored information, not in the way the information is stored or maintained. The notebook contributes fixed content; it does not itself generate graded, revisable, or confidence weighted representations. This point can be maintained even if we grant that the relevant cognitive system is not Otto alone but the coupled system of Otto-plus-notebook (O+N). The issue is not whether interpretation occurs within the system — both biological memory and O+N systems involve processes of interpretation, reconstruction, and error — but where and how those processes are realised. In biological memory, reconstruction is internal to the mechanisms of storage and retrieval themselves: the processes that make recall possible such as selection, abstraction, and contextual modulation, also generate distortion. Error is thus constitutively produced within memory's ordinary operation.

In the O+N system, the notebook contributes stable propositional content, and all reconstruction occurs only at the level of the subject's interaction with that content. The system may misinterpret, misapply, or misremember what is written, but these errors arise from downstream processing applied to a stable informational resource. The result is a division of labour absent in biological memory: storage remains static, while epistemic regulation is imposed upon it. For this reason, even at the level of the coupled system, O+N does not realise a form of memory whose normal operation inherently generates and regulates epistemic risk. A system that retrieves fixed content and then interprets it is not performing the same epistemic labour as

³⁷ Lentoor, Antonio G (2023) 'Cognitive and Neural Mechanisms Underlying False Memories: Misinformation, Distortion or Erroneous Configuration?', *AIMS neuroscience* **10**: 255–68.

one in which retrieval itself is reconstructive, graded, and revisable. The notebook may participate in a cognitive system that errs, but it does not contribute to a system whose errors are generated by the very processes that make remembering possible.

Recent empirical evidence from the field of neuroscience provides further support for this picture of fallible reconstruction and its functional role. Work on fuzzy trace theory, for example, suggests that memory does not store a single, unified representation of past events, but operates through the parallel encoding of two distinct kinds of traces: verbatim traces, which preserve surface-level details, and gist traces, which encode the general meaning or structure of an experience.³⁸ The reconstructive character of remembering depends, in part, on the availability of multiple representational formats. Reyna and Brainerd observe that "subjects encoded more exact representations that... were not used much in reasoning... Thus, a range of representations varying in precision, fuzzy-to-verbatim continua, were available for reasoning."³⁹ While verbatim traces support precise and accurate recall, they are relatively fragile and decay quickly. Gist traces, by contrast, are more durable and play a central role in reasoning, generalisation, and decision-making.⁴⁰ Crucially, however, it is these very gist-based processes that give rise to systematic distortions and false memories. This is not an accidental by-product of an otherwise reliable system, but a consequence of how memory typically functions. In contexts such as survival processing, where information is encoded in a deeper and more meaning-laden way, gist traces are strengthened, leading not only to improved recall but also to an increased likelihood of confidently misremembering related but non-presented information.⁴¹ The same mechanisms that enable abstraction—allowing us to extract patterns, form categories, and apply past experience to new situations—also generate error. Memory's role in learning and adaptation therefore depends on its inherent tendency to misremember; without the ability to move beyond exact detail and represent the general significance of events, cognition would be far less flexible and far less useful.⁴² The challenge for the EMH remains not whether Otto can reliably access information, but whether the notebook participates in the kinds of reconstructive processes that appear constitutive of ordinary memory.

What emerges from this picture of memory is not a passive store of information, but an epistemically regulative system—one that presents remembered information with varying degrees of confidence, salience, and authority. These features are functionally significant as they guide how information is used in reasoning, when it is revised, and how it is integrated with other beliefs. A memory may be tentative or vivid, marginal or central, easily overridden or strongly action-guiding, and these

³⁸ Bialer, Daniel M., Minyu Chang, C J Brainerd, and Valerie F Reyna (2024) 'A Fuzzy Trace Theory Account of Survival Processing', in *Interdisciplinary Perspectives and Advances in Understanding Adaptive Memory*: 402–24. Oxford University Press.

³⁹ Reyna, V F, and C J Brainerd (1995) 'Fuzzy-Trace Theory: An Interim Synthesis', *Learning and Individual Differences* 7: 1–75.

⁴⁰ Bialer et al., *A Fuzzy Trace Theory Account of Survival Processing*, 406.

⁴¹ Bialer et al., *A Fuzzy Trace Theory Account of Survival Processing*, 418.

⁴² Bialer et al., *A Fuzzy Trace Theory Account of Survival Processing*, 422.

modulations are part of what it is to remember at all. The appeal to subjectivity, in this context, is therefore not phenomenological but functional: it concerns the way information is processed and deployed within cognition, rather than how it feels. By contrast, Otto's notebook lacks any such internal structure. It preserves information in a fixed, non-modulating form, and whatever variation in confidence, relevance, or interpretation occurs does so only at the level of Otto's subsequent engagement with it. The notebook does not itself weigh, reinterpret, or reconstruct its contents; it does not generate error through its ordinary operation, but only insofar as it is misread, damaged, or misapplied. For this reason, although the notebook may contribute to a cognitive system that is fallible, it does not realise a form of memory whose characteristic fallibility is constitutive of its function. Otto's notebook can fail as a tool, but it cannot fail in the way a mind does.

§ 5. No Charity from Parity

To see what parity would actually require, we can ask what Otto's notebook would have to be like if it genuinely played the same functional role as biological memory. Reliable storage and accessibility alone would be insufficient. It would need to participate in the same reconstructive, epistemically regulative processes that generate both recall and distortion in ordinary remembering. The entry "MoMA is on 53rd Street" could not simply remain a stable record until deliberately revised. Rather, its content would need to become more or less authoritative, salient, or action-guiding through the ordinary process of retrieval itself, as Otto's circumstances changed. Likewise, revision would need to arise through the same mechanisms that make remembering possible, rather than through the separate act of crossing out an entry, annotating it, or replacing it with a new one. The notebook would therefore need to generate reconstruction as part of its ordinary operation, integrating new contexts into the remembered representation itself rather than preserving fixed content until externally modified.

Otto can, of course, update the notebook when presented with conflicting evidence. However, this requires a distinct act of inscription that produces a new record rather than a transformation of the remembered representation through retrieval itself. The notebook preserves what has been written until it is deliberately altered, and retrieval simply presents that same stored content for subsequent interpretation. Relevant parity therefore requires more than an external record whose contents can be rewritten. It requires a memory system in which retrieval is itself reconstructive: the processes that make remembering possible also integrate contextual information, regulate confidence, and revise the remembered representation. In biological memory, storage, retrieval, and reconstruction are constitutive aspects of a single functional process. An ordinary notebook, by contrast, stores stable propositional content that may subsequently be interpreted, annotated, or overwritten, but those operations are functionally distinct from the retrieval of the stored information itself.

The notebook therefore participates in remembering without realising the same reconstructive organisation that characterises biological memory.

The extended mind hypothesis is compelling in its observation that we think with the world. Tools, symbols, and environments are not peripheral to cognition but are deeply entangled with it, shaping how we reason, remember, and act. C&C are right to resist an unreflective attachment to the boundaries of the skull and to emphasise the continuity between internal processes and our everyday reliance on external resources. However, the move to cognitive status is too quick. The argument from parity depends not simply on the presence of reliable, integrated external resources, but on their realising the same kind of cognitive process as their internal counterparts. It is here that the analogy between Otto and Inga fails. Biological remembering is not a matter of passively retrieving stored information, but of reconstructing the past under conditions of uncertainty, guided by processes that regulate confidence, salience, and revision. Its characteristic fallibility is not a defect to be engineered away, but a feature internal to its function. Otto's notebook cannot participate in this kind of process. However indispensable it may be to his cognitive life, it remains a system that stores and delivers fixed content, leaving all epistemic regulation to the agent who consults it. Its failures arise through breakdown, misinterpretation, or misuse, rather than from the ordinary operation of a reconstructive system. As such, it supports cognition without itself instantiating a memory state. The relevant standard is not whether a resource is used in thinking, but whether it realises the same kind of epistemic work. In the case of Otto and Inga, this standard is not met. If the mind extends, Otto's notebook is not where it ends up.

References

- Adams, Fred, and Kenneth Aizawa (2001) 'Why the Mind Is Still in the Head', in Philip Robbins and Murat Aydede, eds., *The Cambridge Handbook of Situated Cognition*: 78–95. Cambridge University Press
- Bartlett, Frederic C (1995) 'Perceiving, Recognising, Remembering', in *Remembering*: 186–96: Cambridge University Press.
- Bialer, Daniel M., Minyu Chang, C J Brainerd, and Valerie F Reyna (2024) 'A Fuzzy Trace Theory Account of Survival Processing', in *Interdisciplinary Perspectives and Advances in Understanding Adaptive Memory*: 402–24. Oxford University Press.
- Burge, Tyler (1979) 'Individualism and the Mental', *Midwest Studies in Philosophy* 4: 73–121. DOI: 10.1111/j.1475-4975.1979.tb00374.x
- Clark, Andy (2008) *Supersizing the Mind : Embodiment, Action, and Cognitive Extension. Philosophy of Mind*. Oxford University Press.
- Clark, Andy, and David Chalmers (1998) 'The Extended Mind', *Analysis* 58: 7–19. DOI: 10.1111/1467-8284.00096
- Descartes, René (1998) *Discourse on Method and Meditations on First Philosophy*, 4th edition, Donald Cress, trans. Hackett.
- Fodor, Jerry A (1975) *The Language of Thought. Language and Thought Series*. Crowell.
- Lentoor, Antonio G (2023) 'Cognitive and Neural Mechanisms Underlying False Memories: Misinformation, Distortion or Erroneous Configuration?', *AIMS neuroscience* 10: 255–68. DOI: 10.3934/Neuroscience.2023020
- Loftus, Elizabeth F, and Jacqueline E Pickrell (1995) 'The Formation of False Memories', *Psychiatric Annals* 25: 720–25. DOI: 10.3928/0048-5713-19951201-07
- Putnam, Hilary (2011) 'The Meaning of 'Meaning'', in Hilary Putnam *Mind, Language and Reality*: 215–71. Cambridge University Press.
- Reyna, V F, and C J Brainerd (1995) 'Fuzzy-Trace Theory: An Interim Synthesis', *Learning and Individual Differences* 7: 1–75. DOI: 10.1016/1041-6080(95)90031-4
- Schacter, Daniel L (2012) 'Adaptive Constructive Processes and the Future of Memory', *The American Psychologist* 67: 603–13. DOI: 10.1037/a0029869
- Schwitzgebel, Eric (2008) 'The Unreliability of Naive Introspection', *The Philosophical Review* 117: 245–73. DOI: 10.1215/00318108-2007-037
- Shoemaker, Sydney (1963) *Self-Knowledge and Self-Identity* Cornell University Press.
- Sprevak, Mark (2019) 'Extended Cognition', *The Routledge Encyclopedia of Philosophy Online*: 1–15. DOI: 10.4324/9780415249126-V049-1