hundreds and even thousands of pictures and scenes (Delorme, Poncet, & Fabre-Thorpe, 2018; Standing, 1973). It is therefore possible that IAMs and déjà vu experiences differ also in types of representations that get activated. One interesting prediction that emerges from this proposition is that participants scoring high in such experiments may be prone to more frequent self-reported déjà vu.

Financial support. This research received no specific grant from any funding agency, commercial or not-for-profit sectors.

Competing interest. None.

References

Berntsen, D., & Hall, N. M. (2004). The episodic nature of involuntary autobiographical memories. *Memory and Cognition*, *32*, 789–803.

Brewin, C. R. (2014). Episodic memory, perceptual memory, and their interaction: Foundations for a theory of posttraumatic stress disorder. *Psychological Bulletin*, *140*, 69–97. http://dx.doi.org/10.1037/a0033722

Conway, M. A. (2009). Episodic memories. *Neuropsychologia*, *47*, 2305–2313. https://doi.org/10.1016/j.neuropsychologia.2009.02.003

Delorme, A., Poncet, M., & Fabre-Thorpe, M. (2018). Briefly flashed scenes can be stored in long-term memory. *Frontiers in Neuroscience*, *12*, 688. doi:10.3389/fnins.2018.00688

Gardner, R. S., & Ascoli, G. A. (2015). The natural frequency of human prospective memory increases with age. *Psychology and Aging*, *30*, 209–219. https://doi.org/10.1037/a0038876

Mace, J. H. (2004). Involuntary autobiographical memories are highly dependent on abstract cuing: The Proustian view is incorrect. *Applied Cognitive Psychology*, *18*, 893–899. https://doi.org/10.1002/acp.1020

Schlagman, S., & Kvavilashvili, L. (2008). Involuntary autobiographical memories in and outside the laboratory: How different are they from voluntary autobiographical memories? *Memory and Cognition*, *36*, 920–932. doi:10.3758/MC.36.5.920

Schlagman, S., Kvavilashvili, L., & Schulz, J. (2007). Effects of age on involuntary autobiographical memories. In J. H. Mace (Ed.), *Involuntary memory* (pp. 87–112). Blackwell.

Sellen, A., Fogg, A., Aitken, M., Hodges, S., Rother, C., & Wood, K. (2007). Do life-logging technologies support memory for the past? An experimental study using SenseCam. In *Proceedings of the SIGCHI conference on human factors in computing systems* (pp. 81–90). Association for Computing Machinery. doi:10.1145/1240624.1240636

Standing, L. (1973). Learning 10,000 pictures. *Quarterly Journal of Experimental Psychology*, *25*, 207–222. doi:10.1080/14640747308400340

Tulving, E., & Schacter, D. L. (1990). Priming and human memory systems. *Science*, *247*, 301–306. doi:10.1126/science.2296719

Vannucci, M., Pelagatti, C., Hanczakowski, M., Mazzoni, G., & Paccani, C. R. (2015). Why are we not flooded by involuntary autobiographical memories? Few cues are more effective than many. *Psychological Research*, *79*, 1077–1085. https://doi.org/10.1007/s00426-014-0632-y

A rational analysis and computational modeling perspective on IAM and déjà vu

Justin Li[a], Steven Jones[b] and John Laird[b]

[a]Departments of Cognitive Science and Computer Science, Occidental College, Los Angeles, CA, USA and [b]Center for Integrated Cognition, Ann Arbor, MI, USA
justinnhli@oxy.edu
steven.jones@cic.iqmri.org
john.laird@cic.iqmri.org
https://www.oxy.edu/academics/faculty/justin-li
https://scijones.github.io/
https://laird.engin.umich.edu/

doi:10.1017/S0140525X23000092, e367

Abstract

The proposed memory architecture by Barzykowski and Moulin is compelling, and could be improved by incorporating a rational analysis of the functional roles of involuntary autobiographical memory and déjà vu. Additionally, modeling these phenomena computationally would remove ambiguities from the proposal. We provide examples of past work that illustrate how the phenomena may be described more precisely.

The target article by Barzykowski and Moulin (B&M) argues that involuntary autobiographical memory (IAM) and déjà vu are the result of an integrated system of memory and that they naturally arise from recognition and memory retrieval processes. While we agree with this stance, we find the proposed memory architecture lacking in two respects. First, the focus of the authors on the phenomenology of IAM and déjà vu neglects the functional role of these phenomena, the justification for why they might exist, and how they might be used by agents. Second, the memory architecture, as summarized in Figure 1, is only imprecisely specified, leaving room for alternate theories, potential inconsistencies, and omitted details. Here, we consider how performing a rational analysis of, and building computational models of, IAM and déjà vu can mitigate these problems. We use our work on how IAM can support prospective memory as an example of addressing both problems (Li & Laird, 2015), then extend that reasoning to the familiarity judgments that underlie déjà vu.

The rational analysis framework assumes that cognitive processes are optimally adapted to the functional goals of the agent, while subject to ecological constraints and limits on biological and cognitive resources (Anderson, 1990; Lieder & Griffiths, 2020). For memory, we take its primary function to be to "bring past experience to bear on present action" (Anderson, 1994), operating within a small working memory capacity, fixed bandwidth to long-term memory, and other cognitive constraints. This serves as the starting point for understanding the functional role of phenomena such as IAM.

A hypothesis about the structure of memory can be tested via its implementation in a computational model. These models force researchers to be precise in their definitions of the computational representations and processes that underlie their theories and ensure that hypothesized theoretical models of memory are consistent both internally and with broader theories of cognition. This is particularly true in cognitive architectures such as ACT-R (Anderson, 2007), which integrate multiple cognitive processes and potential neural correlates into a single system. This enables the evaluation of their combined performance across multiple tasks, thus ensuring that a hypothesis is compatible with the same mechanisms used to model other phenomena.

Consider, for example, the hypothesis that IAM is the result of automatic matching of sensory and abstract cues with items in memory in order to "quickly rais[e] pertinent information to consciousness without effort" (target article, sect. 4, para. 3). A rational analysis of IAM would start by considering the limits of deliberate retrieval and situations where those limits are exceeded. One such situation is in the prospective memory for future goals, when there may not be the intention to initiate a deliberate retrieval. For example, if one were previously asked to pass a message to a colleague, nothing about seeing the colleague later in the day would necessarily prompt a deliberate retrieval to bring that task to mind, especially if the encounter is otherwise routine. It

is in this context that IAM provides a functional benefit, and indeed, this is known as spontaneous retrieval in the prospective memory literature and is one of several possible strategies for achieving such a goal (McDaniel & Einstein, 2007). This application of rational analysis showcases how IAM can play a role in problem solving: Beyond the passive role suggested in the target article, people can learn to take advantage of IAM to reduce cognitive load.

We have implemented a spontaneous retrieval mechanism in a cognitive architecture, modeled its use in prospective memory, and shown that the conditions under which it succeeds qualitatively resemble results from human experiments (Li & Laird, 2015). More than that, the model requires a fully specified theory of how IAM arises and how it interacts with other memory processes such as deliberate retrieval; in our case, to prioritize problem solving, involuntary retrieval only occurs when no deliberate retrievals are taking place. This decision, which follows from the assumption that memory is used to support goal-driven behavior, suggests an explanation for why researchers have found that IAMs occur most commonly during "relaxed or non-focused state[s] of awareness" and how "being focused would inhibit the activation of knowledge units that are inconsistent with the individual's current goals" (Berntsen, 2008; quoting Mandler, 1994). In this case, the computational model built on the rational analysis framework aligns with the psychology literature.

As for déjà vu, although we know of no existing theory of its functional role in cognition, we agree that it results from false positives in familiarity judgments. Familiarity judgments – or at least its simplest form, recognition – have long been a subject of study via rational analysis, with recognition probability following the optimal Bayesian solution (Shiffrin & Steyvers, 1997). That familiarity is faster than recall allows it to be used to guide the strategic search for knowledge, as per the cognitive-heuristic account of metamemory (Schwartz & Metcalfe, 2011). We have implemented recognition judgments in a cognitive architecture and used it to trigger deliberate retrievals, which led to situations where a false-positive recognition resulted in retrieval failure (Li, Derbinsky, & Laird, 2012). Although we have not modeled déjà vu explicitly, the retrieval failure could suggest that recognition was "implausible." This interaction between deliberate retrieval and plausibility was not explored in the target article, nor how familiarity and recollection interact with each other over time.

In sum, the target article by B&M presents a compelling proposal for how IAM and déjà vu arise. However, their description is missing details that would clarify the relationships between memory mechanisms and could be improved by accounting for the functionality of these phenomena. Applying the rational analysis framework, and considering how the proposed system may be modeled computationally, would resolve these issues.

Financial support. The authors did not receive any outside funding for this work.

Competing interest. None.

References

Anderson, J. R. (1990). *The adaptive character of thought*. Psychology Press.

Anderson, J. R. (2007). *How can the human mind occur in the physical universe?* Oxford University Press.

Anderson, N. H. (1994). Functional memory. In *A functional theory of cognition* (pp. 363–394). Psychology Press.

Berntsen, D. (2008). Involuntary autobiographical memories: Speculations, findings, and an attempt to integrate them. In J. H. Mace (Ed.), *Involuntary memory* (pp. 20–49). Blackwell Publishing. https://doi.org/10.1002/9780470774069.ch2

Li, J., Derbinsky, N., & Laird, J. E. (2012). Functional interactions between memory and recognition judgments. In *Proceedings of the 26th AAAI conference on artificial intelligence* (pp. 228–234). AAAI Press.

Li, J., & Laird, J. E. (2015). Spontaneous retrieval for prospective memory: Effects of encoding specificity and retention interval. In *Proceedings of the 13th international conference on cognitive modeling* (pp. 142–147). University of Gronigen Press.

Lieder, F., & Griffiths, T. L. (2020). Resource-rational analysis: Understanding human cognition as the optimal use of limited computational resources. *Behavioral and Brain Sciences*, *43*, e1. https://doi.org/10.1017/S0140525X1900061X

Mandler, G. (1994). Hypermnesia, incubation, and mind popping: On remembering without really trying. In C. Umiltà & M. Moscovitch (Eds.), *Attention and performance 15: Conscious and nonconscious information processing* (pp. 3–33). The MIT Press.

McDaniel, M. A., & Einstein, G. O. (2007). *Prospective memory: An overview and synthesis of an emerging field*. Sage.

Schwartz, B. L., & Metcalfe, J. (2011). Tip-of-the-tongue (TOT) states: Retrieval, behavior, and experience. *Memory & Cognition*, *39*(5), 737–749. https://doi.org/10.3758/s13421-010-0066-8

Shiffrin, R. M., & Steyvers, M. (1997). A model for recognition memory: REM – retrieving effectively from memory. *Psychonomic Bulletin and Review*, *4*(2), 145–166.

Are involuntary autobiographical memory and déjà vu cognitive failures?

John H. Mace

Department of Psychology, Eastern Illinois University, Charleston, IL, USA
jhmace@eiu.edu
https://www.eiu.edu/psych/

doi:10.1017/S0140525X23000213, e368

Abstract

This commentary supports Barzykowski and Moulin's model, but departs from it on the question of functionality, where IAMs and déjà vu fractionate. The authors seem to say that IAMs are functional, while déjà vu is not. As there is no hard evidence supporting the idea that IAMs are functional, I argue that both phenomena should be viewed as cognitive failures.

Barzykowski and Moulin (B&M) present an interesting model for understanding involuntary autobiographical memories (IAMs) and déjà vu, two mysterious cognitive phenomena. They argue that déjà vu is the likely non-functional by-product (or "side effect") of other cognitive processes, and while they make similar arguments for IAMs, they seem to indicate that IAMs have, in contrast, many functions (see Table 1 in their article). It is at this point, the question of adaptive functionality (e.g., Baddeley, 1988; Bluck & Alea, 2002), that I depart from their model. It is not clear to me how one (déjà vu) is seen as a cognitive failure and the other (IAM) is not. There is plenty of evidence that IAMs are by-products of other processes (e.g., unique cuing and priming; Ball, 2015; Berntsen, Staugaard, & Sørensen, 2013; Johannessen & Berntsen, 2010; Mace, 2005; Mace & Kruchten, 2022; Mace, McQueen, Hayslett, Staley, & Welch, 2019; Mace & Unlu, 2020, see target article). It is also reasonable to assume, even without this evidence, that IAMs, like déjà vu, may merely be cognitive failures. Thus, my commentary focuses on the idea that IAMs, like déjà vu and other similar phenomena, may be cognitive failures.

Proponents of the view that IAMs are functional have a number of challenges to overcome. For example, questionnaire studies on