

Alma Mater Studiorum Università di Bologna  
Archivio istituzionale della ricerca

Déjà vu: A botched memory operation, illegitimate to start with

This is the final peer-reviewed author's accepted manuscript (postprint) of the following publication:

*Published Version:*

Stendardi D, B.A. (2023). Déjà vu: A botched memory operation, illegitimate to start with. BEHAVIORAL AND BRAIN SCIENCES, 46, e378-e378 [10.1017/S0140525X2300016X].

*Availability:*

This version is available at: <https://hdl.handle.net/11585/960901> since: 2024-02-23

*Published:*

DOI: <http://doi.org/10.1017/S0140525X2300016X>

*Terms of use:*

Some rights reserved. The terms and conditions for the reuse of this version of the manuscript are specified in the publishing policy. For all terms of use and more information see the publisher's website.

This item was downloaded from IRIS Università di Bologna (<https://cris.unibo.it/>).  
When citing, please refer to the published version.

(Article begins on next page)

Target Article: Are Involuntary Autobiographical Memory and Déjà Vu Natural Products of Memory Retrieval?

**Déjà vu: a botched memory operation, illegitimate to start with**

Debora Stendardi<sup>1</sup>, Anindita Basu<sup>2</sup>, Alessandro Treves<sup>2</sup>, Elisa Ciaramelli<sup>1</sup>

1.Dipartimento di Psicologia 'Renzo Canestrari', viale Berti-Pichat 5, 40126 Bologna,  
Italy

2.SISSA, Cognitive Neuroscience, via Bonomea 265, 34136 Trieste, Italy

Debora Stendardi: debora.stendardi2@unibo.it

Anindita Basu, abasu@sissa.it

Alessandro Treves, ale@sissa.it, tel: +39 403787623

Elisa Ciaramelli, elisa.ciaramelli@unibo.it, tel: +39 0512091838

## **Abstract**

Rather than a natural product, a computational analysis leads us to characterize déjà vu as a failure of memory retrieval, linked to the activation in neocortex of familiar items from a compositional memory, in the absence of hippocampal input, and to a misappropriation by the self of what is of others.

Abstract wordcount 51

Main text wordcount 988

References wordcount 195

Entire text wordcount 1298

Déjà vu: a botched memory operation, illegitimate to start with

Freud (1901) had already noted that déjà-vu involves memory retrieval: “the uncanny feeling we have, in certain situations, of having had exactly the same experience once before, or of having once before been exactly in the same place, though our efforts never succeed in remembering the previous occasion that announces itself this way.” Barzykowski and Moulin (2022)’s analytical effort conceptualizes déjà-vu as one possible output of a continuum (although structured as a discrete decision tree in Figure 1 of their article) of spontaneous memory retrieval phenomena: while involuntary memories would be the unexpected retrieval of content, déjà-vu would be that of an unjustified feeling of familiarity (not accompanied by a particular memory in mind). Déjà-vu, indeed, soon reveals itself as the phantom of a memory, one we do not belong to, a ‘feeling of retrieval’ to say it with the authors, but one that people describe as disquieting, eerie, awkward (Brown, 2003).

What exactly happens during déjà-vu that disorients us? How can a spontaneous retrieval process go awry? According to the authors, déjà-vu happens when there is a feeling of familiarity that does not pass a plausibility check (Figure 1 of their article), and it is on the neural mechanism of this feeling of *familiarity* and the question of its *implausibility (memory absence)* that our commentary is focused.

Many times we experience familiarity of unclear origin without having a déjà-vu. The butcher-on-the-bus phenomenon, where someone feels familiar but it is not clear from where, is one such case, but it is not disquieting an experience. On the other hand, sometimes déjà-vu occurs in familiar places or involves familiar people, where feelings

of familiarity would be plausible and justified. Moreover, it is not clear how the dichotomous (explicit?) plausibility signal in Figure 1 can be computed in the absence of access to content.

We think that the feeling of familiarity that accompanies the experience of déjà-vu (troubles us because it) is fundamentally different from other instances of familiarity mentioned in the paper: it is not relative to a single item, but to a composition of items, to an experience, albeit fragmented: the place, who was there, some words we uttered, something that will happen next. Hence, we expect the ensuing recollection of the corresponding event that instead does not happen. Perhaps that feels implausible: to have forgotten an entire event that we are currently reliving.

Our recent modelling study (Ryom et al., 2022) offers a computational explanation of associative retrieval failures. These are in fact very frequent, especially if retrieval is triggered by the activation of partial cues in the neocortex, rather than by hippocampal activity indexing memory. Our model network is comprised of ‘Potts units’, which represent patches of cortex, interacting through long-range connections (Figure A). A compositional memory, such as the memory for a complex event (e.g., my dog hid my friend’s sweater in the park), is conceived as composed of several items, each of which has a pre-established neocortical representation (dog, park, sweater). Storing this new memory only involves acquiring the novel connections among participating items. Memory retrieval could be triggered either by the activation of a partial cue in the cortex, which is a variable fraction of the units active in the memory (e.g., sweater + friend), or by a hippocampal input that sustainedly cues all the memory units simultaneously, working as an index to the compressed representation of the entire

memory (see Figure A). One main finding of our study is that the cortical storage capacity for compositional memories is much lower than previously calculated for unitary representations (Treves & Rolls, 1994). The reason is that while the hippocampus is thought to store newly assembled compressed representations of each episode in memory, the neocortex has to make do with reusing pre-established representations of the various components of the episode (Ciaramelli et al., 2006). The ability of the neocortical network to retrieve compositional memories from partial cues, in the absence of hippocampal input, is shown, analytically and with computer simulations, to be severely limited, plagued by the interference from competing representations (Ryom et al., 2022).

On this view, déjà-vu could be characterized as an ‘incomplete’ memory state where some familiar items from a compositional memory (or from several distinct memories) get activated in neocortex (e.g., kids + bench; Figure 1), in the absence of hippocampal input. This activation is sufficient to trigger familiarity for an experience, but not the reinstatement of a full-fledged memory (assuming one exists). The ensuing feeling of familiarity may be particularly uncanny if the partial cue activates self-relevant items or schemata in the neocortex (Stendardi et al., 2021), conferring self-relevance to a memory that might potentially be false, and should last until activated memory fragments are enough to finally trigger monitoring mechanisms that explicitly refute the participation of the self. Or, we suggest, the (false) memory can be abandoned based on an implicit network signal that automatically reads out high levels of simultaneous activity in the neocortex that in the absence of hippocampal activity are more compatible with imagination than with memory. By contrast, the protracted failure of memory monitoring may lead to confabulation, the false memory for unhappened

events (Gilboa et al., 2006; see also Moulin, 2013). Similar to déjà-vu, confabulation entails fragments of memory traces, and is mostly self-related (Gilboa et al., 2006; see also Moulin, 2013). Unlike déjà-vu, confabulation is not abandoned but endorsed confidently. Interestingly, confabulation is triggered by familiar stimuli (Ciaramelli, 2008), and dampened by reducing the cognitive resources available for assembling (wrongly) memory elements (Ciaramelli et al., 2009).

Does the activation of multiple (self-relevant) memory fragments make déjà vu so unique and distinguishable from other illusory familiarity phenomena? Is the estranging feeling associated with déjà-vu the by-product of a just foiled risk of confabulation? Future studies should test this hypothesis, for example studying whether déjà-vu is associated with the activation of ventral prefrontal cortex regions, and the computational conditions conducive to memory or confabulatory signals.

Insert Figure A here

### **Conflict of interest statement**

The authors declare no competing interests.

### **Funding**

The authors acknowledge the support of a PRIN grant from the Italian Ministry of Education, University, and Research to EC and AT (PRIN #20174TPEFJ).

## References

- Brown, A.S. (2003). A review of the déjà vu experience. *Psychological Bulletin*, 129, 394–413.
- Ciaramelli, E. (2008). The role of ventromedial prefrontal cortex in navigation: a case of impaired wayfinding and rehabilitation. *Neuropsychologia*, 46, 2099-105.
- Ciaramelli, E., Gheiti, S., & Borsotti, M. (2009). Divided attention during retrieval suppresses false recognition in confabulation. *Cortex*, 45, 141-53.
- Ciaramelli, E., Lauro-Grotto, R., & Treves, A. (2006). Dissociating episodic from semantic access mode by mutual information measures: evidence from aging and Alzheimer's disease. *Journal of Physiology-Paris*, 100, 142-53.
- Gilboa, A., Alain, C., Stuss, D.T., Melo, B., Miller, S., & Moscovitch, M. (2006). Mechanisms of spontaneous confabulations: a strategic retrieval account. *Brain*, 129, 1399–1414.
- Moulin, C. J. (2013). Disordered recognition memory: recollective confabulation. *Cortex*, 49, 1541-1552.
- Ryom, K. I., Stendardi, D., Ciaramelli, E., & Treves, A. (2022). Computational constraints on the associative recall of spatial scenes. *bioArXiv*.
- Stendardi, D., Biscotto, F., Bertossi, E., & Ciaramelli, E. (2021). Present and future self in memory: the role of vmPFC in the self-reference effect. *Social Cognitive and Affective Neuroscience*, 16, 1205-1213.
- Treves, A., & Rolls, E. T. (1994). Computational analysis of the role of the hippocampus in memory. *Hippocampus*, 4, 374-391.



## Figure A

Figure caption. The hippocampus activates all the five items constituting the real event 'my dog hid my friend's sweater in the park' (straight gray arrows). The activation of two highly familiar items in the absence of hippocampal input may result in déjà vu (fragmented red arrows). Each item has a sparse distributed but partially localized representation over the cortex.

