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

Behavioral and Brain Sciences

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

--Manuscript Draft--

| | |
|---|---|
| **Manuscript Number:** | |
| **Full Title:** | Déjà vu: a botched memory operation, illegitimate to start with |
| **Short Title:** | compositional memory retrieval and déjà vu |
| **Article Type:** | Open Peer Commentary |
| **Corresponding Author:** | elisa ciaramelli<br>University of Bologna Department of Psychology: Universita degli Studi di Bologna Dipartimento di Psicologia Renzo Canestrari<br>ITALY |
| **Corresponding Author Secondary Information:** | |
| **Corresponding Author's Institution:** | University of Bologna Department of Psychology: Universita degli Studi di Bologna Dipartimento di Psicologia Renzo Canestrari |
| **Corresponding Author's Secondary Institution:** | |
| **First Author:** | debora stendardi |
| **First Author Secondary Information:** | |
| **Order of Authors:** | debora stendardi |
| | anindita basu |
| | alessandro treves |
| | elisa ciaramelli |
| **Order of Authors Secondary Information:** | |
| **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. |

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[1], Anindita Basu[2], Alessandro Treves[2], Elisa Ciaramelli[1]

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 834

References wordcount 195

Entire text wordcount 1208

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.” Déjà-vu however soon reveals as the phantom of a memory, one we do not belong to, just a ‘feeling of retrieval’ to say it with Barzykowski and Moulin, which people describe as disquieting, eerie, awkward (Brown, 2003). What exactly happens during déjà-vu that disorients us? How can a spontaneous memory 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. Many other times, however, we experience familiarity of unclear origin. 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. The feeling of familiarity that accompanies the experience of déjà-vu troubles us because 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. And 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, and shows that 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 1). 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 1).

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). Particularly low levels of retrieval success were obtained activating frequent (as opposed to rare) items, which are shared among multiple memories and likely to trigger the retrieval of multiple compatible/interfering memories (e.g., park).

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. On the contrary, the protracted failure of memory monitoring may lead to confabulation, the false memory for unhappened events, which indeed, similar to déjà-vu, 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 promoted by processing 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.

Insert Figure 1 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., Ghetti, 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 1**

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 deja vu (fragmented red arrows). Each item has a sparse distributed but partially localized representation over the cortex.