

False memories for predictable words are not a consequence of prediction but backward semantic association: Evidence from German

Katja Haeuser, Saarland University
khaeuser@coli.uni-saarland.de

Language comprehension is predictive in nature [1,2]. Previous research has suggested that words that are likely to be predicted during reading but are disconfirmed (e.g., “injury” when reading “At the hospital the nurse patches up the pants”), are not suppressed immediately [3,4]. Instead, they remain activated in memory: In downstream recognition memory tests, even those that are administered with a considerable timely delay, predictable disconfirmed words (henceforth, “lures”) often elicit higher rates of false remembering than do unrelated words [5,6]. This has been interpreted as evidence that the pre-activation of the predictable word extends beyond the immediate sentence context, and affects longer-term memory structures. However, it is currently unclear if these false memory effects are really a consequence of pre-activation during sentence processing. False memories may also arise during the recognition memory test when participants use the semantic features of the test word (e.g., “injury”) to re-instate their memory of the previously read sentences, irrespective of any prior pre-activation (e.g., “That word reminds me of that one sentence where a nurse was doing something in a hospital, so I guess that I recognize this word”). Here, we investigate this hypothesis by comparing false memory rates following two critical encoding conditions: One in which the critical word (“injury”) is predictable (“At the hospital the nurse patches up the **pants**”; sentence-final position), and one in which it is not (“The **pants** were patched up by the nurse at the hospital”, sentence-initial position). If false memory rates really are a consequence of prediction during sentence reading, false recognition judgments should be higher for lures in the sentence-final position, where they are predictable. Alternatively, if false memory rates are a consequence of backward semantic association, false recognition judgments should be equally high for sentence-final and sentence-initial lures.

Ninety-three native speakers of German (44m, 48f, 1n-b, *Mean Age* = 28 years, *Range* = 19-40) participated in the study. Participants encoded critical sentences (see Table 1) in a self-paced manner. After a 20minute retention interval (individual-difference tests, not reported on here), they completed the (surprise) recognition memory test. Participants were asked to indicate whether presented nouns were “old” (i.e., previously seen during SPR, $n=44$) or “new” (i.e., previously not seen and not expected, $n=44$), and gave their confidence for each response (high vs low). People also performed recognition judgments for critical lures (i.e., sentence-final and sentence-initial lures, $n=44$).

Hit rates in the recognition memory test were significantly higher than false alarm rates to new items ($p < .001$), leading to an average d' of 1.13 (Range: -0.39 – 3.76), a comparable rate to previous studies (e.g., [3, 7]). Figure 1 shows the effects of condition and confidence on “old” judgments. We analyzed the data by means of a binomial linear mixed effects model (glmer). The model estimated trial-by-trial “old” judgments depending on condition and confidence, including their interaction. According to the model, lures created high levels of false remembering, especially in high-confidence judgments where lures were false alarmed to even more than in low-confidence judgments. Critically though, false alarm rates for lures did not differ between sentence-final and sentence-initial lures, neither as a main effect ($p = .5$), nor in interaction with confidence ($p = 1$). For correct recognition memory (hits), predictable words showed a memory advantage over unpredictable ones ($p = .01$).

Our results call into question that false remembering of predictable words emerges exclusively as a consequence of predictive processing. The false memory effect in the present data is best explained by a backward search strategy that operates during the recognition memory test, where participants are trying to “jog” their memory by generating semantic associations between test words and previously read sentences. This sort of backward search strategy may result in increased hit rates for old, predictable words, as these have stronger semantic associations with the sentence context. Altogether, false remembering is likely driven by prediction *and* backward semantic association.

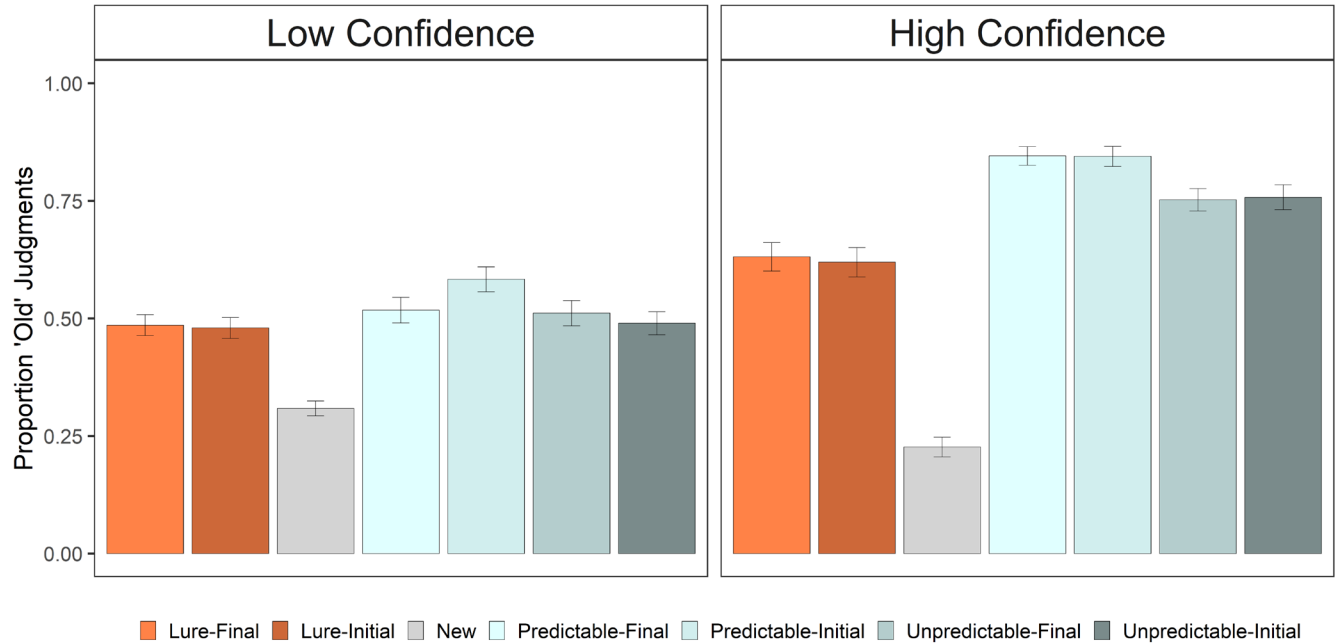
Table 1

English approximations of German experimental sentences used in the reading phase. Bold words were probed as “old” (previously seen) in the recognition memory test. Words in parentheses were probed as lures (i.e., Lure-Final and Lure-Initial).

Predictable-Final	At the hospital, the nurse patches up the injury of the little girl.
Unpredictable-Final	At the hospital, the nurse patches up the pants (<i>injury</i>) of the little girl.
Predictable-Initial	The injury of the little girl was patched up by the nurse at the hospital.
Unpredictable-Initial	The pants (<i>injury</i>) of the little girl were patched up by the nurse at the hospital.

Figure 1

Data plot illustrating the effects of condition and confidence on proportions of “old” judgments in the recognition memory test.



Note. For lures and new words (bars 1-3 in each panel), “old” judgments reflect false alarms. For previously seen words (bars 4-7 in each panel) “old” judgments reflects hits.

References

[1] Kuperberg, G. R., & Jaeger, T. F. (2016). What do we mean by prediction in language comprehension?. *Language, Cognition and Neuroscience*, 31(1), 32-59.

[2] Ryskin, R., & Nieuwland, M. S. (2023). Prediction during language comprehension: What is next? *Trends in Cognitive Sciences*, 27(11), 1032-1052.

[3] Rommers, J., & Federmeier, K. D. (2018a). Lingering expectations: A pseudo-repetition effect for words previously expected but not presented. *NeuroImage*, 183, 263-272.

[4] Rich, S., & Harris, J. (2021). Unexpected guests: When disconfirmed predictions linger. In *Proceedings of the 43rd Annual Meeting of the Cognitive Science Society*. Austin, TX: Cognitive Science Society.

[5] Hubbard, R. J., Rommers, J., Jacobs, C. L., & Federmeier, K. D. (2019). Downstream behavioral and electrophysiological consequences of word prediction on recognition memory. *Frontiers in Human Neuroscience*, 13, 291.

[6] Haeuser, K. I., & Kray, J. (2022). Uninvited and unwanted: False memories for words predicted but not seen. In J. Culbertson, A. Perfors, H. Rabagliati & V. Ramenzoni (Eds.), *Proceedings of the 44th Annual Meeting of the Cognitive Science Society* (pp. 2401-2408).

[7] Rommers, J., & Federmeier, K. D. (2018). Predictability's aftermath: Downstream consequences of word predictability as revealed by repetition effects. *Cortex*, 101, 16-30.