Prediction error may not boost recognition memory for unexpected words, but drives short-term adjustment of semantic networks

Sutong Duan (Sutong.Duan@warwick.ac.uk), Matthew Mak, Chiara Gambi
Department of Psychology, University of Warwick

Comprehenders routinely predict upcoming linguistic input1. While correct predictions facilitate language processing2, the consequences of incorrect predictions, which generate prediction error (PE), remain unclear. Some studies show enhanced memory for plausible but unexpected words3 compared to expected words, others report null4 or even reversed5 effects. While these studies focus on the recognition of individual words, few studies investigated the broader memory consequences of incorrect predictions. If PE drives learning6, its effects should extend beyond item memory to updating relationships between predicted and encountered input. Specifically, prediction errors may promote associative links between expected words and the unexpected words that replace them.

In three experiments, we examined how PE influences memory for unexpected words, their perceived relation to originally predicted words, and the longevity of these effects. The experiments employed a two-phase design, with no delay (Exp 1a and 1b) or a 24-hour delay (Exp 2) between training and test phases. During training, participants heard either sentences that encouraged strong expectations for a target word (e.g., *gun* after "*In the shooting range, without authorisation, people are not allowed to hold a…")* or sentences that were weakly constraining (e.g., *"She stood up to stretch, then bent over and took out a…"*); instead of the predicted word (lure: *gun*), they encountered an unexpected alternative (e.g., *cigarette*), eliciting either higher or lower PE. The subsequent post-context provided scaffolding for the integration of the unexpected word (e.g., "with visible smoke and a glowing red tip"). Memory for the items and their perceived relation was tested using word recognition and semantic relatedness judgment tasks, respectively.

Exp 1a revealed no recognition advantage for unexpected words that elicited higher PE (Figure A). In a close replication (**Exp 1b**), a small effect of PE emerged; combined analyses of Exp 1a and 1b confirmed PE confers at most a small advantage in unexpected word recognition (β =0.29, p=.02). In contrast, in the relatedness judgement task, participants in Exp 1a were more likely to judge *lure–unexpected* pairs (e.g., *gun–cigarette*) as semantically related in the high-PE condition than in the low-PE condition, suggesting the formation of stronger associative links between unexpected words and previously predicted (but unseen) words (Figure B). This effect was further qualified by an interaction with sentence plausibility, such that the difference between high- and low-PE conditions increased as sentence plausibility increased. **Exp 1b** did not find evidence for a main effect of PE, but it replicated the interaction between PE and plausibility, and the robustness of this interaction was further confirmed in a combined analysis of the no-delay experiments (1a and 1b) (β =0.50, p < .001) (Figure C). No effects of PE were found in either task after a 24h delay (Exp 2), suggesting they are short-lived.

In sum, we found only weak and inconsistent (across replications) evidence that PE enhanced recognition for unexpected words, but our findings show, for the first time, that PE strengthens associative links between predicted but unseen (lures) and unexpected words. This effect was modulated by sentence plausibility, indicating that integration of the unexpected word into preceding context is necessary to observe it. It was short-lived, suggesting repeated encounters in similar context may be needed for long-term changes to semantic representations.

References

[1] Pickering, M. J., & Gambi, C. (2018). *Psychological Bulletin,144*(10), 1002–1044. https://doi.org/10.1037/bul0000158

[2] Staub, A. (2015). Language and Linguistics Compass, 9(8), 311-327. https://doi.org/10.1111/lnc3.12151

[3] Hubbard, R. J., & Federmeier, K. D. (2024). *Journal of Cognitive Neuroscience, 36*(1), 1–23. https://doi.org/10.1162/jocn_a_02078

[4] Hubbard RJ, Rommers J, Jacobs CL and Federmeier KD (2019). Front. Hum. Neurosci. 13:291. doi: 10.3389/fnhum.2019.00291

[5] Haeuser, K. I., & Kray, J. (2023). *Language, Cognition and Neuroscience, 38*(4), 558–574. https://doi.org/10.1080/23273798.2021.1924387

[6] Chang, F., Dell, G. S., & Bock, K. (2006). *Psychological Review, 113*(2), 234–272. https://doi.org/10.1037/0033-295X.113.2.234

[7] Mak, M. H. C., Ball, L. V., O'Hagan, A., Walsh, C. R., & Gaskell, M. G. (2025). *Cognition*, *258*, 106086. https://doi.org/10.1016/j.cognition.2025.106086

[8] Haeuser, K. I., & Kray, J. (2022). *Applied Psycholinguistics, 43*(5), 1193–1220. https://doi.org/10.1017/S0142716422000364

Figures