Counterfactual imagination as a source of memory distortion: cognitive and brain mechanisms

Chloe Brunskill[1], Phot Dhammapeera[2], Robin Hellerstedt[3] and Zara Bergström[1]

[1] University of Kent [2] Chulalongkorn University, Thailand [3] University of Cambridge

Background

Previous research where participants were asked to perform or imagine actions with everyday objects have shown that people have difficulty differentiating between imagined and perceived events in their long-term memory(1). However, these studies do not tell us how counterfactual-imagination can distort *existing* memories.

In our study, participants performed actions with everyday objects and later mentally rehearsed the same action or imagined a counterfactual action when shown pictures of the objects. Objects not shown in this task were treated as the baseline condition. We then tested participants memory for the performed actions and recorded EEG activity to investigate the brain mechanisms associated with imagination-induced memory distortions. We predicted that performed actions would be most difficult to recall after counterfactual imagination.

EEG method and analyses:

- EEG was recorded from 27 scalp electrodes plus eye-channels and mastoids, bandpass filtered between 0.1-40Hz, and referenced to bilateral mastoids. Morlet wavelet time-frequency transformation was used to estimate oscillatory power, normalised to a dB scale against a baseline between -.625 to -.375ms.
- For cluster-based permutation tests, t-tests were conducted on every possible timepoint, sample and frequency. T-values of significant clusters were then grouped in space, time and frequency into a summed cluster-level t-value and were then tested against a null-distribution of cluster-statistics created using 5000 random permutations.

Stimuli, procedure and design:

– Rehearsed
– Imagined
– Baseline

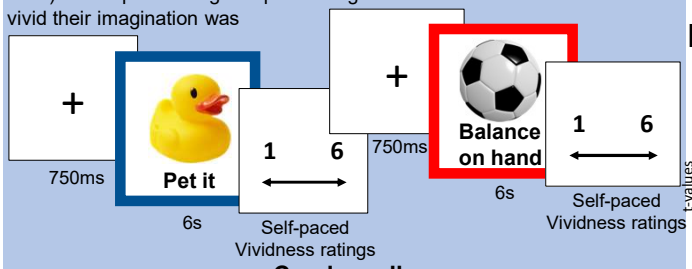
36 participants, 20.2 years, SD_{age} = 1.76 years

Encoding phase

120 real objects handed to participants with instructions to perform actions

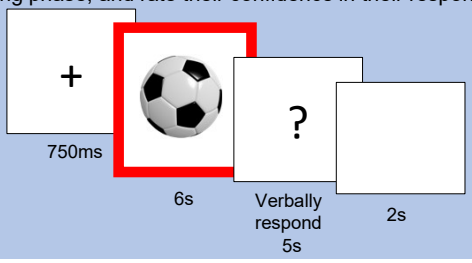
Manipulation phase

80 object picture -action sentence pairs shown with 3 repetitions (240 trials). Participants imagined performing the described action and rated how vivid their imagination was



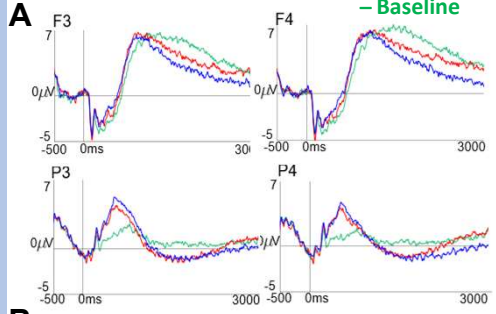
Cued recall

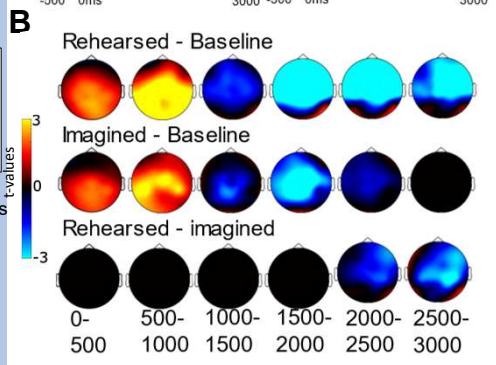
Participants asked to recall the action performed for each object in the encoding phase, and rate their confidence in their response.



Cued recall ERPs

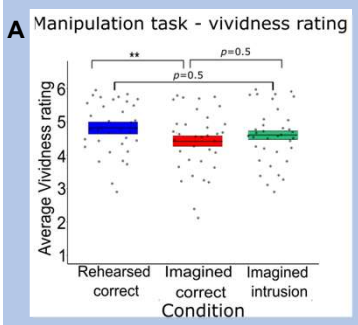
– Rehearsed
– Imagined
– Baseline

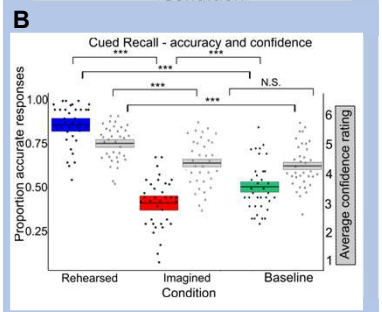


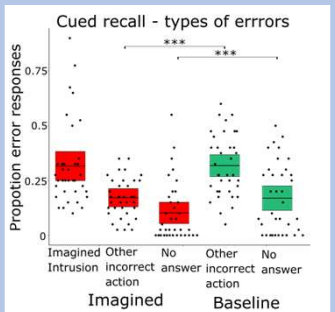


A. Grand-average ERPs from the cued recall test, showing left-frontal (F3), right-frontal (F4), left-parietal (P3) and right-parietal (P4) electrodes.
B. The topography of significant pairwise differences between ERP conditions. Coloured regions show ERP differences that were significant against the cluster-corrected threshold.

Behavioural results

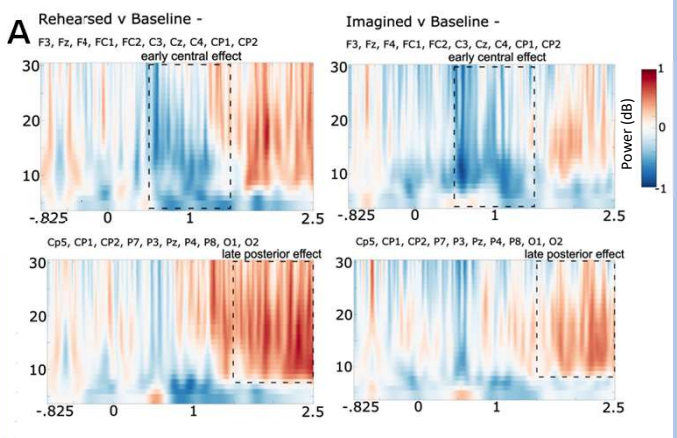


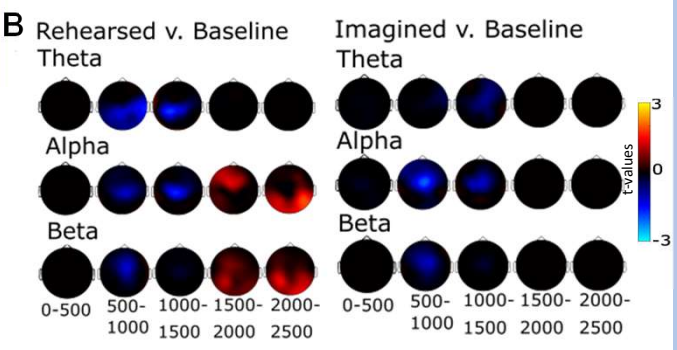




A. Vividness ratings during the manipulation phase, separated by subsequent cued recall accuracy
B. Proportion accuracy and confidence ratings during the cued recall test
C. Proportion of different error types for the imagined and baseline conditions on the cued recall test. .

Cued recall time frequency analysis





A. Grand-average power differences at central (top) and posterior (bottom) electrodes, for rehearsed - baseline (left) and imagined - baseline (right) conditions.
B. The topography of significant pairwise differences between conditions. Coloured regions show power differences that were significant against the cluster-corrected threshold, across Theta (4-7Hz) Alpha (8-12Hz) and Beta (13-30Hz) frequency bands.

Discussion

As predicted, cued recall accuracy was impaired after counterfactual imagination.

Early ERPs (~200-1000ms) were sensitive to whether the image cue had been previously seen, in line with typical ERP old/new effects (showing evidence of cue familiarity and recollection, e.g. (2,3)). Later fronto-central ERP slow drifts (1500-3000ms) were more positive when recall was difficult (baseline > imagined > rehearsed), in line with literature associating frontal slow drifts with executive control functions during remembering, such as retrieval monitoring(4).

EEG oscillations showed a similar pattern, with initial effects tracking memory retrieval success and latter effects tracking the engagement of sustained control processes. In the first half of the time window (~500-1500ms; during the time period when cortical memory reinstatement occurs(5)), rehearsed and imagined conditions were associated with power reductions (event-related desynchronization, ERD), relative to the baseline condition across all frequencies (4-30Hz) with a central topography. Similar ERD effects in the alpha/beta bands have previously been suggested to index recall success (6,7).

A later (~1500-2500ms) alpha and beta (8-30Hz) ERD effect across the posterior scalp was specific for the baseline compared to rehearsed condition. Imagination ERD in this time-window was intermediate. This alpha/beta ERD effect was therefore functionally similar to the late ERP slow-drift, and may also reflect sustained cognitive control/ effort when retrieval required sustained attempts and/or monitoring.

Conclusions:
We found novel evidence that counterfactual imagination can distort true memories of self-performed actions, despite their sensorimotor rich nature. Similar retrieval control processes seemed to be involved when recall was difficult due to a lack of rehearsal (baseline) and when recall was difficult due to the impairing effects of counterfactual imagination.

References:
1. Brandt, V. C., Bergström, Z. M., Buda, M., Henson, R. N., & Simons, J. S. (2014). Did I turn off the gas? Reality monitoring of everyday actions. *Cognitive, Affective, & Behavioral Neuroscience*, *14*(1), 209-219.6.
2. Rugg, M. D., & Curran, T. (2007). Event-related potentials and recognition memory. *Trends in Cognitive Sciences*, *11(6)*, 251-257.
3. Duarte, A., Ranganath, C., Winward, L., Hayward, D., & Knight, R. T. (2004). Dissociable neural correlates for familiarity and recollection during the encoding and retrieval of pictures. *Cognitive Brain Research*, *18*(3), 255-272.
4. Hayama, H. R., Johnson, J. D., & Rugg, M. D. (2008). The relationship between the right frontal old/new ERP effect and post-retrieval monitoring: specific or non-specific?. *Neuropsychologia*, *46*(5), 1211-1223.
5. Staresina, B. P., & Wimber, M. (2019). A neural chronometry of memory recall. *Trends in Cognitive Sciences*, *23(12)*, 1071-1085.
6. Martín-Buro, M. C., Wimber, M., Henson, R. N., & Staresina, B. P. (2020). Alpha rhythms reveal when and where item and associative memories are retrieved. *Journal of Neuroscience*, *40(12)*, 2510-2518.
7. Hanslmayr, S., Staresina, B. P., & Bowman, H. (2016). Oscillations and episodic memory: addressing the synchronization/desynchronization conundrum. *Trends in Neurosciences*, *39*(1), 16-25.