

No evidence of fast mapping in healthy adults using an implicit memory measure: failures to replicate the lexical competition results of Coutanche and Thompson-Schill (2014)

Kshipra Gurunandan, Elisa Cooper, Roni Tibon, Richard N. Henson & Andrea Greve

To cite this article: Kshipra Gurunandan, Elisa Cooper, Roni Tibon, Richard N. Henson & Andrea Greve (2023) No evidence of fast mapping in healthy adults using an implicit memory measure: failures to replicate the lexical competition results of Coutanche and Thompson-Schill (2014), *Memory*, 31:10, 1320-1339, DOI: [10.1080/09658211.2023.2262188](https://doi.org/10.1080/09658211.2023.2262188)

To link to this article: <https://doi.org/10.1080/09658211.2023.2262188>



© 2023 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group



[View supplementary material](#)



Published online: 21 Nov 2023.



[Submit your article to this journal](#)



Article views: 514



[View related articles](#)



[View Crossmark data](#)



This article has been awarded the Centre for Open Science 'Open Data' badge.



This article has been awarded the Centre for Open Science 'Open Materials' badge.



This article has been awarded the Centre for Open Science 'Preregistered' badge.

No evidence of fast mapping in healthy adults using an implicit memory measure: failures to replicate the lexical competition results of Coutanche and Thompson-Schill (2014)

Kshipra Gurunandan^{a,b}, Elisa Cooper^a, Roni Tibon^{a,c}, Richard N. Henson^{a,d} and Andrea Greve^a

^aMedical Research Council Cognition and Brain Sciences Unit, University of Cambridge, Cambridge, UK; ^bBasque Center on Cognition, Brain and Language, Donostia-San Sebastián, Spain; ^cSchool of Psychology, University of Nottingham, Nottingham, UK; ^dDepartment of Psychiatry, University of Cambridge, Cambridge, UK

ABSTRACT

Fast mapping (FM) is a hypothetical, incidental learning process that allows rapid acquisition of new words. Using an implicit reaction time measure in a FM paradigm, Coutanche and Thompson-Schill (Coutanche, M. N., & Thompson-Schill, S. L. (2014). Fast mapping rapidly integrates information into existing memory networks. *Journal of Experimental Psychology: General*, 143(6), 2296–2303. <https://doi.org/10.1037/xge000020>) showed evidence of lexical competition within 10 min of non-words being learned as names of unknown items, consistent with same-day lexicalisation. Here, Experiment 1 was a methodological replication ($N = 28/\text{group}$) that found no evidence of this RT competition effect. Instead, a post-hoc analysis suggested evidence of semantic priming. Experiment 2 ($N = 60/\text{group}$, online study, pre-registered on OSF) tested whether semantic priming remained when making the stimulus set fully counterbalanced. No evidence for either lexical competition nor semantic priming was detected. Experiment 3 ($n = 64$, online study, pre-registered on OSF) tested whether referent (atypicality boosted lexical competition (Coutanche, M. N., & Koch, G. E. (2017). Variation across individuals and items determine learning outcomes from fast mapping. *Neuropsychologia*, 106, 187–193. <https://doi.org/10.1016/j.neuropsychologia.2017.09.029>), but again no evidence of lexical competition was observed, and Bayes Factors for the data combined across all three experiments supported the hypothesis that there is no effect of lexical competition under FM conditions. These results, together with our previous work, question whether fast mapping exists in healthy adults, at least using this specific FM paradigm.

ARTICLE HISTORY

Received 20 August 2022
Accepted 19 September 2023

KEYWORDS

Fast mapping; episodic encoding; implicit memory; lexical integration; word learning

Fast mapping (FM) was a concept originally introduced to explain rapid vocabulary acquisition in infants (e.g., Carey, 1978; Carey & Bartlett, 1978). While the concept is still a matter of debate (see Carey, 2010; O'Connor et al., 2019; O'Connor & Riggs, 2019), there has been additional recent debate about whether FM also exists in adults (see Cooper et al., 2019a, and associated commentaries). This debate was triggered largely by a specific FM paradigm introduced by Sharon et al. (2011), which was designed to resemble infant vocabulary learning, with incidental learning of a new item via a deductive inference in the presence of a known item. More specifically, participants in this paradigm are presented with two pictures, one depicting an unfamiliar item and the other depicting a semantically-related known item, and asked a yes/no

question concerning the name of the unfamiliar item. To answer the question, participants must deduce that the name applies to the unfamiliar item by comparing the two items (see ahead to Figure 1 for an example). Participants are typically not informed that their memory for the name of the unfamiliar item will be tested later. In a control condition – the episodic encoding (EE) condition – participants are presented with a single unfamiliar item and its name, and instructed simply to remember the name. In both conditions, after a delay (e.g., 10 min or 24 h), participants perform a three-alternative-forced-choice (3AFC) recognition memory task, which requires them to choose a picture that matches the name (a test of “explicit”, or “declarative”, memory; Schacter & Tulving, 1994). Sharon et al. (2011) ran this paradigm on

CONTACT Richard Henson  rik.henson@mrc-cbu.cam.ac.uk  Medical Research Council Cognition and Brain Sciences Unit, University of Cambridge, 15 Chaucer Road Cambridge CB2 7EF, UK

 Supplemental data for this article can be accessed online at <https://doi.org/10.1080/09658211.2023.2262188>.

© 2023 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. The terms on which this article has been published allow the posting of the Accepted Manuscript in a repository by the author(s) or with their consent.

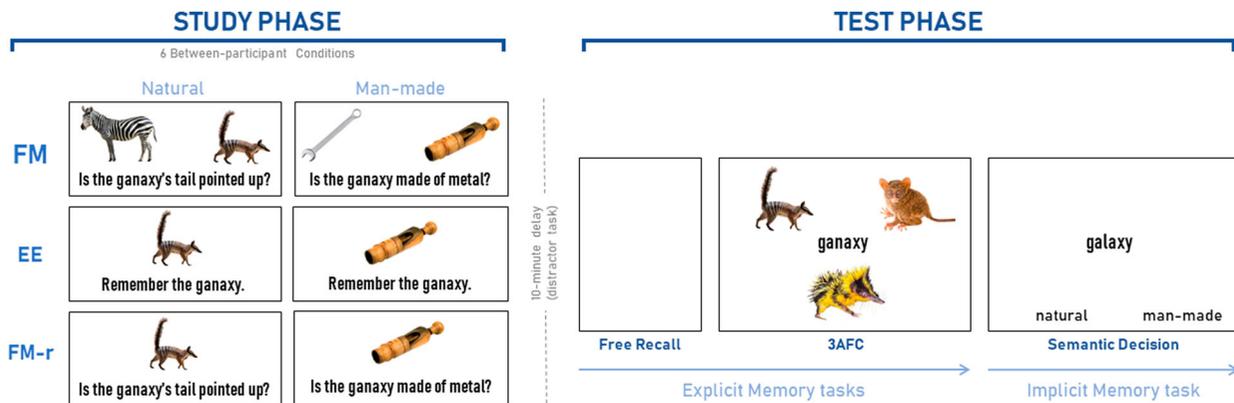


Figure 1. Experimental procedure for all three experiments. The study phase consisted of six between-participant conditions: Experiment 1 tested conditions FM (fast mapping) and EE (explicit encoding) with natural stimuli, Experiment 2 tested conditions FM, EE and FM-r (FM without referent) with both natural and man-made stimuli, and Experiment 3 tested only the FM condition with both natural and man-made stimuli, and varied the typicality of the referents within-participant. Assignment of stimuli to condition was counterbalanced across participants within each experiment. In FM and FM-r study conditions, names were to be incidentally associated with the unknown pictures. Key prompts for “yes”/“no” were displayed at the bottom of the screen on respective sides, which have been omitted from the figure for simplicity. In the EE study condition, participants were instructed to learn the unknown object’s name. The study phase was followed by a 6–10 min delay task, and then the test phase. The test phase was identical across conditions and experiments, and involved two explicit memory tasks, followed by an implicit memory test. The explicit tests started with free recall of the new names, and then a three Alternative Force Choice (3AFC) test containing three possible pictures. In the implicit test, participants made a speeded, semantic category judgement about hermit words, half of which had studied neighbours and the other half did not (key prompts were displayed at the bottom of the screen, counterbalanced across participants). For more information, see main text.

healthy adults and four patients with amnesia following medial temporal lobe damage. Their findings were striking: whereas patients performed worse than controls on the EE condition, as expected given their amnesia, they performed as well as controls on the FM condition, across both 10-minute and 24-hour retention intervals. In other words, the FM paradigm had recovered the patients’ memory performance to that of controls.

This finding is not only practically important, e.g., for rehabilitation, but also theoretically important. The latter is because most neuroscientific theories of declarative memory propose that the hippocampus is initially responsible for storing new information, and only after a period of consolidation does this responsibility shift to the cortex for longer-term retention (McKenzie & Eichenbaum, 2011; Norman & O’Reilly, 2003; Squire & Bayley, 2007). Because the hippocampus was damaged in Sharon et al.’s patients, their intact performance in the FM condition, even after 10 min, suggests that sometimes new declarative information can be encoded in the cortex directly (and retrieved explicitly). This led to the concept of fast cortical mapping (FCM).

Subsequent to the Sharon et al.’s (2011) study, there have been several attempts to replicate the FCM benefit for individuals with amnesia. An experiment from the same lab as Sharon et al. (2011), but with three new patients (Experiment 2 in Merhav et al., 2014), replicated the patients’ intact memory performance in the FM condition and impaired performance in the EE condition, but only when there was no interference between object-names (FM performance collapsed if objects were repaired with new names). However, other studies failed to replicate Sharon et al.’s results in adults with acquired amnesia (Cooper et al., 2019a; Smith et al., 2014), or in

individuals with developmental amnesia (Elward et al., 2019); nor was evidence of an FM benefit found in adults with amnesia when using a different FM procedure (Warren et al., 2016; Warren & Duff, 2014).

Using healthy participants, we previously found no interaction between Sharon et al.’s FM and EE conditions in older relative to younger participants, despite MRI-confirmation of smaller hippocampal volumes in the older group (Greve et al., 2014). However, other studies with healthy controls have found that various manipulations affect FM and EE performance differently, such as the level of interference (Merhav et al., 2014), effects of sleep (Himmer et al., 2017), and retention interval and prior knowledge (Li et al., 2020). Nonetheless, one result that has been found consistently across all studies is that performance in the FM condition never exceeds that in the EE condition, i.e., there is currently no evidence of FM providing a memory advantage in healthy populations using explicit memory tests (see also Cooper et al., 2019b).¹ Sharon et al. (2011) explained this by an intact, hippocampal episodic memory system over-shadowing any fast cortical learning.

However, there is one situation where a memory advantage of FM over EE has been reported in healthy adults, namely when using implicit tests of memory, rather than the explicit tests used in all the studies cited above. The neural mechanisms that support implicit (non-declarative) memory, like priming, are thought not to require the hippocampus, even for new associations like object-names (e.g., Goshen-Gottstein et al., 2000). In two experiments using an implicit measure based on response times (RTs) to make a semantic decision, Coutanche and Thompson-Schill (2014) reported evidence of learning in FM but not EE conditions. Their RT measure was claimed to index

“lexical competition”, in terms of slower response times to known words when a new competitor (the new name) had been learned. Based on the previous logic of Bowers et al. (2005), the names of the to-be-learned unfamiliar animals (e.g., “ganaxy”) were lexical neighbours of real hermit words (e.g., “galaxy”). Hermit words are those that have no orthographic neighbours, and cannot be made into another English word through letter substitution or deletion. It is well-established that the greater the neighbourhood density of a word, the longer the RT to recognise that word, due to competition from similar word forms (Andrews, 1996; Bowers et al., 2005; Davis & Taft, 2005). Therefore, if a hermit word’s neighbourhood is increased from none to one through learning a neighbour as an animal “name”, then RTs for that word should be slower than for hermit words that have not had a neighbour studied.

Integration of a new word into the lexicon, such that it can affect recognition of its neighbours, is often thought to require a period of consolidation, particularly sleep (except under specific circumstances involving repeated exposures; see Kapnoula et al., 2015; Lindsay & Gaskell, 2013). Yet Coutanche and Thompson-Schill (2014) reported a lexical competition effect only 10 min after performing the FM condition. More specifically, RTs to make a man-made/natural decision to hermit words (e.g., “galaxy”) were significantly slower when a new (non-word) neighbour (e.g., “ganaxy”) has been inferred as the name of an unfamiliar animal during the FM study phase, relative to hermit words with no studied neighbour. They also tested explicit memory and found the opposite pattern to the implicit test: Explicit memory was better for the new animal names under EE than FM, consistent with all other studies above, and suggesting an important dissociation between implicit and explicit measures.

The same group later published boundary conditions for the implicit competition effect under FM (Coutanche & Koch, 2017). Firstly, they reported that the competition effect was only seen in participants with high “semantic trait score” (using an equal split of their participants). Secondly, they reported that the competition effect increased with the atypicality of the semantic referent object (for its taxonomic category; see also Coutanche, 2019); indeed, for typical referents, they reported the opposite effect, i.e., faster rather than slower RTs for studied words relative to unstudied words. However, this makes it unclear why they previously found a competition effect when collapsing over all participants (without knowledge of their semantic trait score) and all trials (regardless of referent typicality) in their two original experiments (Coutanche & Thompson-Schill, 2014).

After completion of our Experiments 1 and 2 (the latter pre-registered), another study (Zaiser et al., 2021) reported a partial replication of the lexical competition effect of Coutanche and Thompson-Schill (2014). In Experiment 1

of their study, Zaiser and colleagues report lexical competition (i.e., slowing for hermit words with neighbours) four minutes after FM encoding, when combining their “high” and “low” feature overlap conditions (which refer to the degree of feature overlap between known and unknown items). However, they did not run a control condition, such as an explicit encoding (EE) condition, to test whether this lexical competition effect is specific to the FM task. They did include an EE condition in their Experiments 2–4, but also switched to a semantic priming test, where primes were the previously studied names of novel objects, and targets were real words whose semantic category was either the same or different to that of the object named by the prime. When this test was performed a few minutes after study, the authors found a priming effect for FM in the high feature overlap condition, but not in the EE condition or low overlap FM condition. There were no priming effects at the 24-hour test.

However, it is unclear whether this semantic priming effect reflects the same lexicalisation process indexed by the RT competition effect of Coutanche and Thompson-Schill (2014). For instance, the observed semantic priming effect in the FM condition could potentially be attributed to the rapid retrieval of the studied referent object and the associated category, which facilitates the semantic decision for related words while inhibiting decisions for unrelated words. This could explain why the effect is not present when there is no referent in the EE condition, or when the referent is likely to belong to a different category in Zaiser et al.’s low overlap condition. Interestingly, in the intentional FM condition, where a referent object was also presented, no semantic priming effect was found. It is possible that the intentional learning strategies employed by participants in this condition may have diverted their attention from processing the referent object in sufficient detail (unlike the incidental condition where processing of the referent is naturally evoked). Furthermore, another study (McGregor et al., 2020) failed to replicate the FM advantage of Coutanche and Thompson-Schill (2014): though these authors focused on explicit tests in people with developmental language disorders versus those with typical language development. In their Supplementary Materials, they report that they did not find evidence of same-day lexical learning in an implicit memory test.

In summary, the original finding from Coutanche and Thompson-Schill (2014) is theoretically important because it supports prior claims that the FM paradigm of Sharon et al. (2011) engages a distinct learning process in healthy adults, and also because it is one of the few situations in which same-day lexicalisation has been claimed. However, there is some uncertainty in the literature about whether the FM paradigm does engage a distinct learning process, and no study has reported a complete replication of Coutanche and Thompson-Schill’s (2014) results, whereby an implicit

lexical competition effect (slowing of RTs to hermit words for which a neighbour has been studied) is found after an FM task but not EE task. We ran three experiments to attempt this replication.

Experiment 1

Methods

The current experiment is identical to Experiment 1 of Coutanche and Thompson-Schill (2014), with four minor exceptions: (1) Only the 10 min delay condition was used, because the objective here was to investigate the FM lexicalisation effect within the same day; (2) a semantic priming test was not included because Coutanche and Thompson-Schill reported no significant effect after 10 min; (3) the “names” are from Coutanche and Thompson-Schill’s Experiment 2, because this allowed a greater number of stimuli in order to match their counterbalancing across EE and FM conditions and across hermit words with and without studied neighbours; (4) the pictures were changed to reflect UK norms (as in Greve et al., 2014; see Stimuli and Design section).

The experiment consisted of two learning conditions (FM and EE). As in Coutanche and Thompson-Schill (2014) these were manipulated between participants (in order to minimise contamination of the FM condition by explicit encoding strategies, as might happen if learning condition was manipulated within participants; see Cooper et al., 2019a). Regardless of learning condition, the test phase was identical: explicit memory tests of Free Recall and 3AFC recognition, followed by an implicit memory test using semantic decision (the “lexical integration task” of Coutanche & Thompson-Schill, 2014).

Participants

Fifty-six young volunteers (aged 19–41, mean 24.6 years, 38 females) were recruited from the MRC Cognition and Brain Sciences Unit’s Volunteer Panel and compensated financially for their time.

Volunteers were randomly allocated ($N=28$ in each group; 18 females in one group and 20 in the other) to one of two learning conditions (either the FM or EE). An N of 28 per group was chosen to provide full counterbalancing (see Stimuli section below) and provide a power of over 80%. Power was calculated from Coutanche and Thompson-Schill’s reported effect size of Cohen’s $d=0.69$ in the same day condition of their Experiment 1, when comparing the size of the competition effect across their EE and FM groups. Using GPower 3.1.9.2 (Faul et al., 2007), our N of 28 per group provides power of 81.7% for detecting an effect this big using a one-tailed, unpaired t-test (for a greater competition effect in the FM than EE group) with an alpha of .05. We note however that the effect size of Coutanche and Thompson-Schill may have been inflated by publication bias, and that one would generally need 2.5

times their sample size (i.e., 62 per group) to have an 80% chance of concluding that the effect is not “undetectably small” (Simonsohn, 2015), which would correspond to a Cohen’s $d=0.35$ for 33% power with Coutanche and Thompson-Schill’s $N=25$ sample size. We address this later when combining data across our experiments.

The EE and FM groups did not significantly differ (t ’s < 0.81 , p ’s > 0.42) in age, years of education, and general intelligence “ g ” as measured by the Cattell Cultural Fair Scale 2 intelligence test (Cattell & Cattell, 1960). Though participants were randomly allocated to learning condition, the EE group had higher Verbal IQ than the FM group, 117 versus 111, respectively, $t(54)=2.02$, $p=0.048$, two-tailed, as measured by the Wechsler Abbreviated Scale of Intelligence (WASI) (Wechsler, 1999) (which was the filler task). All participants were native English speakers, had normal or corrected-to-normal vision, and provided informed consent prior to their participation. Their inclusion was approved by the Cambridge Psychological Research Ethics Committee (reference 2005.08).

Stimuli

In keeping with Coutanche and Thompson-Schill (2014), only pictures of animals were used. The pictures were 64 colour photographs: 32 of rarely-known but real animals, and 32 of known animals from the same category as the unknown animals. All were a subset of the culturally-normed “unknown” animals from previous published sources (Cooper et al., 2019b; Greve et al., 2014).

Unknown animal “names” came from two lists of 16 neighbours of hermit words (e.g., “ganaxy” as a neighbour for “galaxy”) from Experiment 2 in Coutanche and Thompson-Schill (2014). Unknown animal pictures were randomly paired with the hermit word neighbours, which were kept in their original lists as published by Coutanche and Thompson-Schill, creating two sets of 16 picture-word pairings, such that each set could appear in either the FM or EE group. Pictures of known animals were used in the FM study section as the semantic referent; their names were never used. Only the unknown 16 pictures and names from the study phase were presented in the 3AFC recognition memory test. All 32 real English hermit words (e.g., galaxy) were used in the implicit memory test. The assignment of set to studied versus unstudied was counterbalanced within and across groups.

Procedure

The paradigm is shown in the top of Figure 1. E-prime 2.0 (Psychological Software Tools, Pittsburgh, PA, 2012) was used to display stimuli and collect button press responses. Participants completed a Study phase and a Test phase. Participants were randomly assigned to either the FM condition or the EE condition, in which they either incidentally or explicitly, respectively, learned the names (hermit word neighbours) of unknown animals. There was a 10-minute

delay, filled with a verbal task (see below), between Study and Test phases. In the Test phase, which was identical regardless of the Study Phase, explicit memory for the animal names was first tested by free recall and then by 3AFC. Then, implicit memory was tested using a task methodologically identical to Coutanche and Thompson-Schill's "lexical integration" task. Hermit words with studied neighbours and those without were presented for a speeded man-made versus natural semantic decision. Response keys for yes/no (FM study phase) and natural/man-made were counterbalanced across participants. When response side was combined with set (Set 1 and Set 2) and condition (EE and FM), there were 8 unique counterbalancings. Participants sat 26 in. in front of a LCD 19 inch computer monitor. All items were displayed on the screen using a white background; text and fixation crosses were black 24 point Arial font, except for the 3AFC test where text was 28 point font.

Study phase. In the FM condition, participants were presented with background "ruse" information; they were informed that it was a visual object perception experiment but not that their memory would be later tested. In the EE condition, participants were instructed to remember the animals and names for a later memory test. Both conditions started with a white screen for 500 ms, followed by the stimuli, which were displayed for six seconds, and this cycle continued until all 16 unknown stimuli had been viewed twice. The list was randomly presented once before being presented again in a different random order. In the FM condition, a known animal picture (e.g., dog) and an unknown animal picture (e.g., numbat, named "ganaxy") were displayed in the centre of the screen with a yes/no question below ("Is the ganaxy's fur striped?"), such that each unknown animal was seen once on each side (central-left or central-right) and once with each response type ("yes" or "no"). Prompts for key assignments appeared in the lower-left and lower-right corners of the screen for the respective response. Stimuli remained on the screen for six seconds regardless of when the response was given. If no response was given within the six seconds stimuli display, then the stimuli disappeared and an on-screen prompt to respond was displayed until response. The known animal was from the same category as the unknown animal, and the question referred to the unknown animal and a feature present in only one animal (e.g., striped fur, though fur would be present in both animals). For each of their two presentations, unknown animals were paired with a different known animal and a different orienting question. In the EE condition, an unknown animal was displayed centrally, with its name and instructions to remember it below ("Remember the naskin."); this was identical for both presentations, and no key press was required. Prior to the Study phase, participants completed a separate run of 10 practice Study trials with feedback and unique stimuli.

Test phase. The study phase and test phase were separated by a 10-minute delay during which participants performed a verbal task to prevent rehearsal, which was the WASI (Wechsler, 1999). Vocabulary and Similarities sub-tests were used to compute a Verbal IQ score, in keeping with the procedure of Coutanche and Thompson-Schill (2014) who report using an "unrelated vocabulary test and questionnaire". In the test phase, memory for the unknown animal names from the study phase was measured with a Recall memory test, a 3AFC recognition memory test and an implicit memory test, as expanded below.

In the Recall test, participants were given a maximum of two minutes to report aloud as many new names from the Study phase as possible, which were noted by the experimenter. In the 3AFC test, each of the 16 studied "names" was presented centrally together with two other studied unknown pictures, one at the top-left, one at the top-right and one at the central-bottom of the screen. Participants indicated which picture matched the name by using one of three keys, with each of the three locations corresponding to the correct response approximately equally often. Each unknown animal picture was displayed three times, once in each location, and once as the correct answer and twice as a foil. Trials were displayed in a random order, remained on the screen until response or a maximum of six seconds, and were separated by a 500 ms white screen, in keeping with Coutanche and Thompson-Schill (2014).

In the implicit memory task, the 16 hermit words whose neighbours were studied and the 16 hermit words whose neighbours were not studied were displayed one-at-a-time in a random order in the centre of the screen for 500 ms. Immediately, prior to each word, a white screen was presented for 350 ms, preceded by a central fixation cross for 800 ms. Participants were asked to respond as quickly and as accurately as possible, via left or right keys on the keyboard, whether the word represented a man-made or natural item. Prompts for key assignments appeared in the lower-left and lower-right corners of the screen, respectively. If a response was not made during the word display, then an additional maximum two seconds was given for a response with only the key prompts displayed. Feedback about whether or not the response was correct appeared for one second following the response. Participants first completed five practice unique trials that lead seamlessly into the real test.

After the Test phase, participants completed a familiarity test, in which they were shown each unknown animal picture once more, and reported their familiarity with the item prior to the Experiment using a three-point scale (where 1 = no familiarity). Participants also completed the Cattell culture fair test (Cattell & Cattell, 1960) to measure fluid intelligence ("g").

Analyses

Data were analysed using R, version 3.5.0 (R Core Team, 2020) and RStudio Version 2022.02.3 (RStudio Team,

2020). For all tests, i.e., recall, 3AFC and the implicit memory test, analysis included only nonwords paired with animals that were not pre-experimentally familiar (a rating of 1 on familiarity scale), as reported in the familiarity test (this was also the case for the analyses of Coutanche & Thompson-Schill, 2014). The median number of items reported as pre-experimentally known was 1 (of the 16 items per set; minimum 0, maximum 10 from one participant). These are excluded from all analyses below.

As in Coutanche and Thompson-Schill (2014), the main measure was reaction times (RTs) in the implicit memory test, further restricted to correct responses only, and trimmed to those more than 300 ms and less than 1500 ms. For the main planned comparison, to replicate Coutanche and Thompson-Schill (2014), we used a directional (one-tailed), two-sample t-test to test whether the competition effect (difference in RTs for hermits with versus without a studied neighbour) was larger in the FM Group than EE Group.

We also performed linear mixed effects analysis across individual trials, including both participant and stimulus (word) as random effects. Rather than trimming, here we analysed all correct trials, and used an inverse transform (1/RT) to handle the positive skew in RTs (Ratcliff, 1993). We also added a third within-participant factor of congruency, to look at semantic priming effects (see Results). Although hermit words were either natural or man-made, their studied neighbour (i.e., nonword name) was referred to a natural object. Hence, a natural hermit word was always semantically congruent with its neighbour, whereas a man-made hermit was always incongruent with its neighbour.

For completeness, we also analysed accuracy on both the implicit and explicit memory tests. For the implicit (man-made/natural) task, accuracy was the proportion of trials responded to accurately (natural or man-made; chance was 0.50). Because of ceiling effects, these data were arcsine transformed and then analysed using a mixed-measure ANOVA with between-participant factors of learning condition (FM or EE) and a within-participant factor of hermit word (with or without studied neighbour). In the explicit Recall test, learning was measured as the total number of names correctly recalled aloud. Some participants recalled no words (particularly in the FM Group), resulting in floor effects, and were therefore submitted to a Wilcoxon test. For the 3AFC test, the measure of learning was proportion of correct responses, with chance being 0.33, and compared via one-tailed T-tests.

The R code and data are available in OSF: <https://osf.io/dpvbf/>.

Results

Main planned comparison: competition RTs in Implicit Task

We begin with the results for the main planned comparison for which the experiment was powered, namely whether the competition effect (mean RT for hermits with studied

neighbour minus mean RT for those without) in the implicit memory test was bigger for the FM group than EE group, as Coutanche and Thompson-Schill (2014) reported.

Table 1 (top) shows the means of trimmed RTs for correct trials (as well as mean number of such trials from which the mean was estimated) for each condition and group separately. These are also plotted in Panel D of Figure 2, along with the mean and spread of their differences, i.e., competition effect, in Panel E. For reference, Panels A and B are the corresponding results replotted from Experiment 1 of Coutanche and Thompson-Schill (2014). Most important are the competition effects, which show a competition effect for FM but not EE group of Coutanche and Thompson-Schill (2014) (Panel B), but little evidence of a competition effect for either FM or EE groups in the present Experiment 1 (Panel E).

The key planned, one-tailed, unpaired t-test was designed to test whether the mean competition effect was bigger for the FM group (−5.61 ms) than EE group (−0.21 ms). As expected from Table 1/Figure 2, this effect was not significant, $t(54) = -0.372$, $p = .64$, with an estimated Cohen's d of −0.10. This contrasts with Experiment 1 of Coutanche and Thompson-Schill (2014) for which a significant effect with an estimated Cohen's d of +0.69 was reported.

Anovas and accuracy in implicit task

Even though the interaction is equivalent to the planned comparison performed above, we also subjected the mean trimmed RTs to a 2×2 mixed ANOVA, with between-participant factor of Group (EE/FM) and within-participant factor of Study (studied/unstudied), to test for any main effects. Despite a numerical trend for faster RTs in the FM group, neither the main effect of Group, $F(1,54) = 1.14$, $p = .29$, nor main effect of Study, $F(1,54) = 0.16$, $p = .69$, reached significance.

We also analysed accuracy on the semantic natural/man-made word judgement. Because performance was close to ceiling in both groups (Table 1), an arcsine transform was applied to render the data more Gaussian. There was a numerical trend for higher accuracy in the EE than FM group (see next Section), but in the same ANOVA as above, none of the main effect of Group, $F(1,54) = 2.82$, $p = .10$, main effect of Study, $F(1,54) = 2.45$, $p = .12$, nor most importantly their interaction, $F(1,54) = 0.738$, $p = .39$, reached significance. The latter suggested no speed-accuracy trade-off.

Group differences in verbal fluid intelligence?

A reviewer pointed out that a later study by Coutanche and Koch (2017) failed to find rapid lexical competition after FM in their participants who had semantic memory trait scores that were lower than the median of their sample (see Discussion), and it is possible that our FM group had low semantic trait scores. Our FM group did have significantly lower WASI verbal intelligence scores (VIQ) than our EE group (by chance alone, since group allocation was random), and it is possible that lower semantic

Table 1. Means and standard deviations (in brackets) from implicit tests in Experiment 1 ($N = 56$ total), Experiment 2 ($N = 180$ total) and Experiment 3 ($N = 64$ total), as a function of between-participant learning conditions (groups) and within-participant manipulation of hermit words with or without studied neighbours.

	Implicit memory test					
	Hermits <i>without</i> studied neighbour			Hermits <i>with</i> studied neighbour		
	Accuracy	N trim	Mean RT	Accuracy	N trim	Mean RT
Experiment 1						
EE	0.90 (0.11)	14.18 (1.94)	754.07 (131.36)	0.90 (0.11)	12.79 (2.36)	753.86 (157.32)
FM	0.85 (0.11)	13.50 (1.64)	717.97 (132.17)	0.89 (0.08)	12.71 (2.48)	712.36 (130.82)
Experiment 2						
EE	0.85 (0.14)	13.53 (2.23)	725.54 (133.16)	0.83 (0.13)	12.25 (2.75)	718.60 (140.50)
FM-r	0.83 (0.13)	13.30 (2.09)	754.64 (130.32)	0.82 (0.12)	11.98 (2.54)	745.81 (151.96)
FM	0.83 (0.12)	13.22 (1.91)	714.13 (138.28)	0.81 (0.16)	12.32 (2.50)	718.36 (150.03)
Experiment 3						
FM-atypical	0.83 (0.12)	13.27 (1.91)	822.05 (154.76)	0.80 (0.15)	6.03 (1.39)	826.65 (153.55)
FM-typical				0.85 (0.15)	6.59 (1.29)	799.38 (146.01)

"Accuracy" is the proportion of pre-experimentally unknown items whose man-made/natural judgement was correct, where chance is 0.50. "N trim" mean number of correct trials (max of 16; max of 8 in Experiment 3 condition "with studied neighbours") after removing (trimming) trials with RTs < 300 ms or > 1500 ms, following Coutanche and Thompson-Schill (2014); "Mean RT" is the mean reaction time (RT) for such trials in milliseconds. Experiment 1 had EE and FM groups, Experiment 2 included an additional FM-r group, Experiment 3 examined only the FM group (with referent typicality as a within-participant factor; see Methods). The data for Experiments 2 and 3 are averaged over whether the referent object was natural or man-made.

memory scores correlate with lower verbal intelligence scores (even though the former is more of a measure of crystallised intelligence and the latter is more of a measure of fluid intelligence). This might have reduced our chances of detecting rapid lexical competition in our FM group, and could explain why there was a numerical trend for higher accuracy in the EE than FM group on the semantic decision task.

Firstly, we should point out that, though vIQ was 6 points lower on average in our FM than EE group (111 versus 117), it was still higher than the population average (100). Nevertheless, we fit a linear model that predicted the competition effect as a function of Group, vIQ and their interaction. There was no significant positive effect of vIQ , $t(52) = 0.011$, $p = .99$, and there was still no effect of Group or interaction with vIQ , $F_s(1,52) < 0.27$. Even when run on the FM group alone, there was no sign of a relationship between competition effect and vIQ , $t(52) = 0.84$, $p = .41$. Thus we think it is unlikely that our results were affected by low verbal fluid intelligence. While it is possible that our groups also differed in semantic trait scores, and that these diverged sufficiently from verbal fluid intelligence, we think this is extremely unlikely, and note that the original study on which the present experiment was powered, i.e., Coutanche and Thompson-Schill (2014), did not distinguish participants by semantic trait scores in either of their experiments.

Explicit memory tests: recall and recognition

Given that we found no difference between Groups in the Implicit task, it is important to establish that the two

learning conditions differed in some way. We therefore analysed the data in the explicit memory tests, which we would expect from prior literature (see Introduction) to be higher following EE than FM learning. Recall data were near floor and therefore were analysed using a non-parametric Wilcoxon test. As would be expected when comparing performance on incidental versus intentional learning conditions, participants in the FM condition recalled significantly fewer names (Median = 0.00, IQR = 1.00) than participants in the EE condition (Median = 3.00, IQR = 3.25), $W = 686$, $p < .001$.

The same was found with 3AFC accuracy performance: memory was significantly better in the EE condition ($M = 85.5\%$, $SD = 14.0\%$) than FM condition ($M = 58.6\%$, $SD = 17.9\%$), $t(54) = 6.26$, $p < .001$, with a large Cohen's $d = 1.67$ (though was still above chance (33.3%) in the FM condition, $t(27) = 7.54$, $p < .001$). Indeed, the 3AFC results were numerically very similar to Experiment 1 of Coutanche and Thompson-Schill (2014) – cf. Figure 2C and F – suggesting similar participant populations under this measure at least.

Post-hoc analyses

The "lexical integration" task of Coutanche and Thompson-Schill (2014), which furnished the implicit RT measure analysed above, actually involves a semantic judgement: participants decide whether real hermit words denoted either a natural or a man-made item. Though our data showed no evidence of a competition effect under FM (or EE), the RTs (like those in Coutanche & Thompson-Schill, 2014) might have been confounded by the semantic congruency/incongruency between the

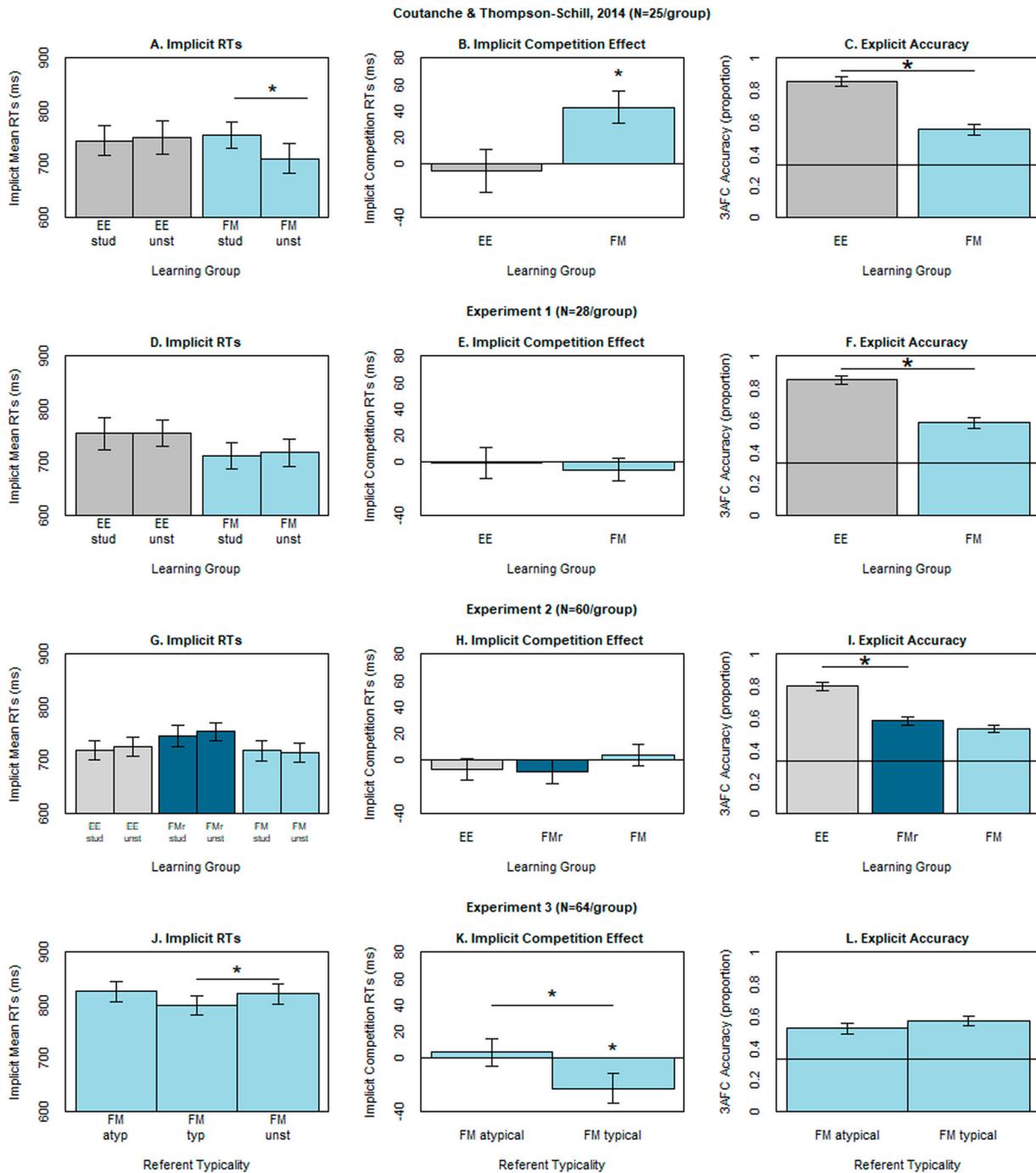


Figure 2. Each row shows data from a separate experiment, the columns contain (from left to right): mean trimmed reaction times (RTs) in the implicit memory test (Panels A, D, G, J) as a function of whether the hermit word had a studied neighbour in each group (EE, grey; FMr, dark blue; or FM, light blue); competition effects after subtracting unstudied from studied RTs (Panels B, E, H, K); three alternative-force-choice (3AFC) accuracy in the explicit test (Panels C, F, I, L). Data in row 1 are re-plotted from Experiment 1 of Coutanche and Thompson-Schill (2014) (C&T-S); data in third and fourth rows are averaged over the referent category (natural/man-made). All RTs are for correct trials only, after excluding pre-experimentally known items. 3AFC chance is marked at .33. * = significant to $p < 0.05$, two-tailed. Error bars are standard error of the mean. ms = milliseconds.

studied neighbour and the judgement. In other words, deciding whether the hermit word was natural or man-made could be affected by the learning of its neighbour as a natural item (i.e., an animal name) in the study phase: this might facilitate RTs for natural hermit words, but inhibit RTs for man-made hermit words. (The 3AFC test where the studied neighbours are presented with

only animal pictures as choices might reinforce this semantic link). This would be a form of semantic priming, possibly related to that reported by Zaiser et al. (2021). With this in mind, we re-examined the data, splitting hermit words according to whether they were semantically congruent with the study category (of natural animals). Of the two hermit-neighbour word lists

of 16 items from Coutanche and Thompson-Schill, one list had six natural items (and ten man-made) and the other had seven natural items (and nine man-made).

For a potentially more sensitive test, we also used a linear mixed effect model on RTs of individual trials, which can accommodate random effects of both participant and stimulus (hermit word). Moreover, rather than trimming RTs like Coutanche and Thompson-Schill (2014), we included all trials with correct responses, but used an inverse transform (1/RT), which has been argued to perform well for the skewed RT distributions in these situations (Ratcliff, 1993). Not only did this recover more trials, but also rendered the distribution more Gaussian than the above trimming. The model predicted (inverse) RTs as a function of all possible interactions between Group, Study and Congruency, together with random intercepts for participant and stimulus.²

Like in the above analysis of mean, trimmed RTs across trials, there was still no evidence for the interaction between Group and Study reported by Coutanche and Thompson-Schill (2014), $F(1,1428.13) = 0.28$, $p = .60$. There was however evidence for a three-way interaction between Group, Study and Congruency, $F(1,1427.37) = 6.30$, $p = .012$ (but no other significant effects, F 's < 1.56, p 's > 0.21). We therefore followed-up the three-way interaction by analysing each Group separately, with random intercept and Study slopes for stimulus, and random intercept and Congruency slopes for participant.

In the FM Group, there was a significant two-way interaction between Study and Congruency, $F(1,32.1) = 5.61$, $p = .024$, but there was no such significant interaction for the EE Group, $F(1,29.5) = 1.36$, $p = .25$. The lack of effect in the EE group could reflect the lack of another natural referent object at study (see Discussion).

To calculate an across-participant effect size for this moderation of the Competition effect by Congruency, we averaged inverse RTs across trials in the FM group, and calculated the competition effect (i.e., subtracted Unstudied from Studied). As expected from the above linear mixed effects model, there was a significant difference between the competition effect for Congruent and Incongruent trials, $t(27) = 2.93$, $p = .007$, two-tailed, Cohen's $d = 0.55$. As shown in Table 2, Congruent trials showed shorter RTs for Studied than Unstudied trials, t

(27) = 2.75, $p = .011$, two-tailed, whereas Incongruent trials showed a trend towards the opposite, $t(27) = 1.65$, $p = .11$, two-tailed. Thus, when collapsing over congruency, as in the previous analyses above, the advantage from congruency and disadvantage from incongruency would tend to push the average competition effect to zero.

Discussion

Though Experiment 1 was a methodological replication of Coutanche and Thompson-Schill (2014), there was no evidence of significant same-day lexical competition (i.e., slowing of RTs to hermit words with a studied neighbour), following either FM or EE. This fails to support the claim by Coutanche and Thompson-Schill that the FM condition (but not EE condition) enables same-day lexicalisation of novel words, i.e., fast mapping.

Nonetheless, in a *post-hoc* analysis, where we split trials according to whether the hermit word was natural versus man-made (which was the dimension probed in the speeded semantic decision task used to assess implicit memory), we did find significant priming effects: RTs for natural hermit words (after FM learning) were speeded when those words had a studied neighbour relative to when they did not, whereas RTs for man-made hermit words were slowed when those words had a studied neighbour relative to when they did not. This pattern is more consistent with a semantic priming effect than a lexical competition effect. This is similar to the semantic priming effect reported by Zaiser et al. (2021), who explored the influence of feature overlap on lexical competition within the FM procedure. Their experimental design involved presenting an unknown item alongside a known referent, with varying degrees of overlap in visual features. Interestingly, they did not find any significant differences in lexical competition between the feature overlap conditions. However, they did report a significant semantic priming effect, which was particularly pronounced when high feature overlap was present.

These findings may initially seem contradictory to the results reported by Coutanche and Koch (2017), who suggested that atypical referents enhance lexical integration. However, it is important to consider that Zaiser et al. manipulated feature overlap independently from typicality, which might tap into distinct underlying processes. In the high overlap condition, the known and unknown items consistently belonged to the same semantic category (e.g., mammal – mammal, vegetable – vegetable), whereas in the low overlap condition, the items belonged to different semantic categories (e.g., mammal – vegetable, bird – mammal). Based on their results, Zaiser et al. proposed that high feature overlap promotes rapid semantic integration, regardless of whether the known item is a typical exemplar of the category. They speculated that increased demands on processing highly similar visual stimuli during FM encoding could enhance semantic integration and the retrieval of associated

Table 2. Means and standard deviations (in brackets) across participants in ms of trial-averaged competition effect (difference in RTs of studied minus unstudied, after reversing inverse transform) from the FM groups in Experiments 1–3, as a function of congruency (whether the hermit word was congruent with the natural/man-made category of the studied item) and the natural/man-made category of the studied items (all of which were natural in Experiment 1, like Coutanche and Thompson-Schill, 2014).

	FM group	Congruent	Incongruent
Experiment 1	Natural	–36.85 (67.17)	+30.25 (73.31)
Experiment 2	Natural	–2.79 (95.12)	+8.21 (64.79)
	Man-made	10.33 (71.10)	–13.29 (93.96)
Experiment 3	Natural	–19.55 (160.76)	+6.64 (73.77)
	Man-made	–35.96 (94.62)	–38.97 (110.30)

meanings. However, it remains to be determined whether the same underlying processes support the semantic priming effect observed in our study.

A key distinction in our experimental design is that we did not manipulate the visual similarity of the referent items. Instead, we observed a semantic priming specifically for test items that belonged to the same semantic category. More specifically, because all studied neighbours were new names of objects that were natural (animals), these RT effects can be explained by congruency/incongruency between the natural/man-made category of the hermit word and the natural category of its new studied neighbour: when they are congruent (e.g., the natural hermit word “galaxy” and animal named “ganaxy”), spreading activation from the studied neighbour to the hermit word could increase the speed of responding “natural” in the implicit test, whereas when they are incongruent (e.g., the man-made hermit word “napkin” and animal named “naskin”), this spread of activation potentially causes interference (slowing). Importantly, it should be noted that this explanation does not necessarily imply that the studied neighbour needs to be represented as a lexical node in the lexicon. Instead, the spread of activation to the hermit word may occur when the neighbour is encountered during study, without that neighbour being explicitly stored in memory. This is a form of semantic priming, rather than lexical competition, and semantic priming effects often occur within the same day (Collins & Loftus, 1975; McNamara, 2005; Tulving & Schacter, 1990). The lack of an overall competition effect (collapsing across hermit words’ semantic category) could be explained by facilitatory semantic priming effects being matched by equally-sized inhibitory semantic priming effects, though this would not explain why Coutanche and Thompson-Schill (2014) found a net inhibitory effect.

It is unclear why such a semantic priming effect would occur only after FM learning, and not EE learning. One possibility is that the presence of a referent object and mention of a semantic feature in the yes/no question in the FM condition (but not EE condition) increased the number of items activated that are from the natural category, e.g., if the hermit word, its studied neighbour and its studied referent were all brought to mind by spreading activation. If so, the effect should be reduced if the referent is removed from the FM condition (similar to Zaiser et al., 2021). We addressed this possibility in Experiment 2.

Experiment 2

In Experiment 2, we attempted again to replicate Coutanche and Thompson-Schill’s (2014) same-day lexicalisation after FM learning, but with the additional factor of semantic congruency. In order to create a fully factorial design, one half of the studied objects were now man-made, to add to the other half that were the natural objects from Experiment 1. This way, we could de-found semantic congruency and semantic category (i.e.,

a natural study object and natural hermit word, or a man-made study object and man-made hermit word would be congruent, whereas a natural study object and man-made hermit word, or a man-made study object and natural hermit word would be incongruent). The semantic priming hypothesis predicts that, when studied neighbours (e.g., “ganaxy”, “naskin”) are names for unfamiliar natural pictures (i.e., animals), semantic decision RTs to natural words (galaxy) will be speeded and RTs to man-made words (napkin) will be slowed. When studied neighbours are unfamiliar man-made objects, RTs for natural words (galaxy) will be slowed and RTs for man-made item words (napkin) will be speeded. That is, there would be an interaction between the congruency of the semantic category of the studied object and the hermit word, and whether or not the item’s neighbour was studied.

We also took the opportunity to add a third learning condition (i.e., test a third group of participants, see Figure 1). This was the “FM-r” variant of FM learning that we previously used in Cooper et al. (2019b) to test whether the presence or absence of the object referent is necessary to obtain a FM effect in explicit memory that is distinct from EE learning (and similar to the variants used by Coutanche & Thompson-Schill, 2014, and Zaiser et al., 2021). Previous studies of fast mapping suggested that a known referent is important (Sharon et al., 2011), e.g., in helping support integration of the new information into existing semantic knowledge (Coutanche & Thompson-Schill, 2015). If so, then any effects of fast mapping seen in the FM condition should not be seen in the FM-r condition, which is identical apart from lacking the simultaneous presentation of a known referent picture beside the unknown item whose name is to be learned. Using tests of explicit memory, Cooper et al. (2019b) found no difference between the FM-r condition and FM condition. However, using the present implicit measure, Coutanche and Thompson-Schill (2014, Experiment 2) found evidence for “lexicalisation” in the FM condition, but not their “IE” condition (the equivalent of our FM-r condition where no known referent picture was shown), supporting a role for the referent in FM. Note they, like us, found no difference between their FM and “IE” conditions when using a test of explicit memory instead. Likewise, Zaiser et al. (2021, Experiments 3–4) found no semantic priming without a reference object in their “IE” condition. Thus, the importance of a referent may again vary according to whether explicit or implicit memory measures are used.

In the present Experiment 2, there were 2×3 between-participant conditions: three groups (FM, FM-r and EE) where the “names” were for unfamiliar (natural) animals (as in Experiment 1, but with the addition of the FM-r condition), and another three groups (FM, FM-r and EE) where “names” were for unfamiliar man-made items. According to Coutanche and Thompson-Schill (2014), a competition effect (slowing of RTs) would be seen only in the FM condition, regardless of whether the study items or hermit

words were natural or man-made. According to the semantic priming effect suggested by our *post-hoc* analysis of Experiment 1, a competition effect (slower RTs) should be found when the natural/man-made status of the studied object is incongruent with that of the hermit word, but there should be a semantic priming effect (faster RTs) when these are congruent. This interaction should also be greater in the FM condition than FM-r or EE condition, if the presence of a referent object of the same natural/man-made category as the studied neighbour increases the amount of semantic priming.

Prior to starting data collection, Experiment 2 was pre-registered on the Open Science Framework (OSF) using a replication registration report template. The pre-registration for conditions FM and EE is located at <https://osf.io/atkp4>; the pre-registration for the FM-r condition is located at <https://osf.io/p7s4f>.

Methods

Experiment 2 was identical to Experiment 1 bar four exceptions: (1) an additional between-participant condition using man-made items as study phase pictures; (2) an additional between-participant FM-r learning condition in the natural and man-made conditions, which was identical to the FM condition, but the known referent picture was not presented; (3) data were collected online; (4) the task separating study-test phases was a letter-digit substitution task rather than the WASI test (so that it could be administered online).

The experiment consisted of six conditions: three learning conditions (FM, FM-r, and EE) crossed with two study categories (natural or man-made items), all manipulated between-participants. Regardless of study phase, each participant completed three types of memory test: Free Recall, 3AFC, and the implicit memory test with semantic decision; the latter is our focus. In this implicit test, there was a within-participant factor of study condition (hermit word with and without a studied neighbour), and a within-participant factor of semantic congruency (whether hermit word category matched that of studied object).

Participants

One-hundred and ninety-three young participants (aged 18–40) were recruited through Prolific (<https://prolific.com/>), which is a web-based crowd-sourcing platform that can be integrated with online experiments. The final dataset was comprised of 180 participants³ from a potential pool of over 3,000 that met the recruitment criteria below. Participants were randomly allocated to one of the six conditions ($N = 30$ in each group: 18 females and mean age 28.3 years in the “FM-natural” condition; 22 females and mean age 29.8 years in the “FM-man-made” condition; 24 females and mean age 29.3 years in the “FM-r-natural” condition; 25 females and mean age 27.8 years in the “FM-r-man-made” condition; 22 females and

mean age 29.03 years in the “EE-natural” condition; 22 females and mean age 28.1 years in the “EE-man-made” condition).

An N of 60 (summing across referent category group) was chosen to provide power of over 80% for detecting a semantic priming effect in the FM group. Using GPower 3.1.9.2 (Faul et al., 2007), the power estimate was 96% given an alpha of .05 and an effect size of Cohen’s $d = 0.55$ from Experiment 1 for a one-tailed, paired t-test for a smaller competition effect for Congruent than Incongruent trials (see <https://osf.io/atkp4>).⁴ For reference, the same N had 98% power to detect the effect size reported by Coutanche and Thompson-Schill (2014) for a one-tailed, unpaired t-test of greater lexical competition in FM than EE conditions.

All participants provided informed consent online prior to the study and were compensated financially for their time. They reported being monolingual English speakers, who were UK citizens currently residing in the UK, since the stimuli were normed for a UK population. Participants also reported having normal or corrected-to-normal vision. They previously participated in a minimum of two online studies on the Prolific platform (<https://prolific.co/>) with an approval rating of at least 95%. The programme of research was reviewed by Cambridge Psychological Research Ethics Committee and received a favourable opinion (reference PRE2016.055). Procedures accorded with the Declaration of Helsinki.

Stimuli

Hermit words and their invented neighbours were identical to Experiment 1. Word-picture stimuli pairs in the unfamiliar “natural” conditions were identical to Experiment 1. The pictures in the unfamiliar man-made conditions were created using identical criteria to the natural conditions, which are detailed in Experiment 1. The 32 pictures of unfamiliar man-made items were a subset of previously published unfamiliar objects (Taylor et al., 2014). Unknown man-made items were randomly paired with the hermit word neighbours, which were kept in their original lists from Experiment 1, creating two sets of 16 picture-word pairings. Pictures of known man-made items from the same semantic category as the unfamiliar items (e.g., car parts, musical instruments) were sourced from the internet to use as the known items in the FM condition. As in Experiment 1, the assignment of stimulus set to with (studied) versus without (not studied) neighbours was counterbalanced across groups, and each set had an equal number of $N = 15$ participants in each condition. Stimuli are available here: <https://osf.io/vjcd6/>.

Procedure

The paradigm is shown in Figure 1. Data collection occurred via the internet on the participants’ computer at a location of the participants’ choice. The experiment was only available via a desktop or laptop computer, and was not compatible with hand-held devices, e.g., phones

or tablets. The experiment was programmed so that it would not proceed unless the display was maximised. Prior to data collection, we pre-specified data replacement criteria. In the implicit memory test in Experiment 1, only one participant had below chance accuracy of 50%. We therefore based our criteria for replacing data on this, and participants with a mean accuracy of less than 50% or RT more than three standard deviations from average were removed and replaced.

The experiment was programmed using a free, open-source tool, JsPsych based in JavaScript (de Leeuw, 2015, <http://www.jspsych.org/>). The MRC-CBU servers hosted the experiment using free, open-source JATOS (Lange et al., 2015, <https://www.jatos.org/>). The servers are based in the EU and are compliant with data protection and security policies.

Study phase. Study phase details are identical to those detailed in Experiment 1. The additional FM-r conditions (FM-r-natural and FM-r-man-made) were identical to the FM learning conditions, except the semantic referent picture was not presented, i.e., only the unfamiliar picture was presented in the center of the screen. The yes/no questions to be answered by the participant were identical to the questions used under the FM learning condition with response prompts, again, on the bottom of the screen. In FM and FM-r conditions, participants were given “ruse” instructions that the task investigated visual perception of pictures and how they should answer questions based on this. They were not told that their memory would be later tested. In the EE condition, participants were instructed to remember item names for a later memory test. Prior to the Study phase, participants completed a separate run of 10 practice Study trials with feedback in the FM and FM-r conditions (no response is given in the EE study phase) and unique stimuli. As in Experiment 1, there was no test phase in the practice.

Test phase. Study and Test phases for each condition were separated by a minimum delay of six minutes, when participants completed three blocks of a non-verbal letter-digit substitution task (an online adaptation of this task: https://healthabc.nia.nih.gov/sites/default/files/dsst_0.pdf). The first block lasted one minute and the two following blocks lasted two and a half minutes each. Each block was preceded by instructions and six practice trials. This distractor task was designed to be administered freely online and unrelated to the main experiment. The test phase methods for all six between-participant conditions were identical to Experiment 1 with only the following exceptions: (1) in the recall memory test, participants reported recollected names from the initial study phase by typing into a response box, which was presented until participants submitted their responses, and (2) in the implicit memory test, practice trials were presented in a separate block prior to an instruction screen announcing the start of the real trials.

Analyses

As in Experiment 1, items that were reported as pre-experimentally familiar were excluded from analyses of Recall, 3AFC, and the semantic decision task. The median number of items reported as pre-experimentally known was 0, ranging up to 11. If we use Coutanche and Thompson-Schill’s (2014) procedure of removing participants with more than half of the stimuli judged as familiar, this would mean removing one participant in the FM-r condition who said they were familiar with 11 of the 16 animals. However, we suspect this high number can be attributed to button mapping error or misinterpreting instructions, so it was assumed they were unfamiliar with all items, as for the median participant. In any case, the FM-r condition was not of primary interest for replication of Coutanche and Thompson-Schill’s (2014) findings regarding the FM and EE conditions.

In order to test the lexicalisation hypothesis of Coutanche and Thompson-Schill (2014), the data were first analysed by collapsing across the two within-participant’s factors, i.e., semantic category of the study item and semantic congruency with the hermit word, in order to match the initial analyses done in Experiment 1 and Coutanche and Thompson-Schill (2014).⁵ As in Coutanche & Thompson-Schill, this analysis was done on the mean of trimmed RTs.

This analysis was then expanded to include the within-participant factor of semantic congruency, to test the post hoc finding from Experiment 1. As in Experiment 1, we transformed RTs rather than trimmed them, though this time the inverse transform was not sufficient to render Gaussian, so a log-inverse transform was used. Finally, the analysis was extended to add the within-participant factor of semantic category (of the unknown object) and expanded to a mixed effect model, in order fit to individual trials, as in Experiment 1.

The R code and data are available in OSF: <https://osf.io/dpvbf/>.

Results

Main planned comparisons: competition RTs in Implicit Task

Though this experiment was powered to detect a modulation of the competition effect by congruency in the FM group (semantic priming in the implicit task), we start by reporting a second test of the proposed greater competition effect in the FM groups than EE groups (averaged across congruency) that was reported by Coutanche and Thompson-Schill (2014) and tested in Experiment 1.

Table 1 (middle) shows the means of trimmed RTs for correct trials (as well as mean number of such trials from which the means were estimated) for FM, FM-r and EE groups, split by whether or not a neighbour was studied, but averaged over natural/man-made referent and

congruency. These are also plotted in Panel G of Figure 2, along with the competition effect (studied – unstudied) in Panel H. There was now a small positive competition effect in the FM group but not EE or FM-r group, but it was much smaller than in Coutanche and Thompson-Schill (2014). Indeed, the planned one-tailed, unpaired *t*-test for whether the mean competition effect was bigger for the FM group (+4.23 ms) than EE group (–6.94 ms) did not reach significance, $t(118) = 0.98$, $p = .16$, with an estimated Cohen's *d* of +0.18.

The second planned comparison for semantic priming within the FM groups also did not reveal a significant effect, i.e., failed to replicate the congruency effect observed in the post hoc analysis of Experiment 1. The one-tailed, paired *t*-test on log-inverse transformed RTs within the FM groups (collapsed over referent category) provided no evidence that the competition effect for Congruent trials ($M = +1.45$ ms) was smaller than for Incongruent trials ($M = -6.22$ ms), $t(59) = -0.73$, $p = .77$, $d = -0.095$.

As in Experiment 1, we also modelled the data using a mixed ANOVA and then linear mixed effects models (see Supplementary Materials). However, despite some numerical trends, we again failed to find any significant effects. Analyses of the accuracy in the three explicit and implicit memory tasks have also been reported in the Supplementary Materials.

Post-Hoc analyses

In a later study, Coutanche and Koch (2017) claimed that the FM lexical competition effect is modulated by the typicality of the referent item, a factor that was not controlled by Coutanche and Thompson-Schill (2014) and hence also not considered in our attempt to replicate their findings. However, it is possible that the typicality of referent items differed between studies, potentially explaining our failure to replicate the 2014 findings. To address this possibility, we obtained typicality ratings for the referent items used in our experiments from an independent group of 27 participants and compared them against the typicality rating means and SDs reported in Coutanche and Koch (2017) for stimuli from both their studies. The results (see Supplementary Material) revealed that the referents in Coutanche and Thompson-Schill (2014) were rated as significantly more typical (by their raters) than were the referents used in our Experiment 1 and Experiment 2 (by our raters⁶). The findings of Coutanche and Koch (2017) suggest that more atypical referents should result in greater lexical competition, hence, it seems unlikely that our failure to replicate owes to referent typicality. Nevertheless, since we could only compare the means of ratings (by different populations in different studies), we could not reliably test whether more atypical referents might have led to greater lexical competition effects. Hence, we investigated this possibility in Experiment 3.

Experiment 3

In this experiment, we attempted again to replicate Coutanche and Thompson-Schill's (2014) same-day lexicalisation after FM learning, this time manipulating the typicality of the referents with which the unknown items were studied. We divided the items into two sets such that one half of the items were studied with typical referents and the other half with atypical referents (as determined by independent typicality ratings). Coutanche and Koch (2017) found a significant main effect of referent typicality on the size of the lexical competition effect, and reasoned that typicality plays a strong role in the neural organisation of concepts in the anterior temporal lobes (which were proposed to underlie fast mapping by Sharon et al., 2011). In a subset of their participants, they showed that atypical referents led to lexical competition while typical referents led to significant facilitation in RTs. However, findings were not reported for their full FM group ($n = 30$), with a second subset exhibiting neither competition nor facilitation effects. Nevertheless, if the findings of Coutanche and Koch (2017) hold, we would expect to see a significant lexical competition effect in at least the atypical condition, hence showing same-day lexical competition effects after FM learning as in Coutanche and Thompson-Schill (2014). This experiment was pre-registered on OSF (<https://osf.io/wqf83>).

Methods

Experiment 3 was identical to Experiment 2 with two exceptions: (1) we only had an FM condition, and (2) we included an additional within-participant factor of Referent Typicality (see Supplementary Material). Hence, the experiment had a single FM learning condition, with a between-participant factor of Category (natural or man-made items), and a within-participant factor of Referent Typicality (atypical or typical referent).

Participants

Seventy adult participants (aged 18–40) were recruited through Prolific. The final dataset was comprised of 64 participants⁷ who were randomly allocated to one of the two between-participant conditions: FM-natural and FM-man-made ($N = 32$ in each group: 14 females and mean age 31.03 years in the “FM-natural” condition; 13 females and mean age 29.19 years in the “FM-man-made” condition).⁸

For the lexical competition effect, we powered this study using the effect size of Cohen's $d = 0.67$ reported by Coutanche and Thompson-Schill (2014) for studied vs. unstudied items in the FM condition. Our sample size of $N = 64$ provides 99.9% power to detect an effect size of this magnitude or greater using a one-tailed, paired-sample *t*-test. However, if the effect is only expected in the atypical condition (all referents in Coutanche and Thompson-Schill's Experiment 1 were mid-typicality), our power might be reduced by virtue of having half as

many trials within participants. Nonetheless, estimating a revised effect size of $0.67/\sqrt{2}=0.48$ instead, our power is still 97.8% (though the precise change in power from reduced trial numbers depends on the relative size of within- versus between-participant variance; Baker et al., 2021). To test for semantic priming, we averaged the effect sizes of $d=0.55$ and $d=-0.095$ from our Experiments 1 and 2 respectively, i.e., $d=0.23$, which results in 55% power to detect the effect with a one-tailed, paired-sample *t*-test. Consequently, we might not have power to detect such a small semantic priming effect if present, but note that semantic priming is only of secondary interest of this study, for which we might not be able to make strong conclusions.

All participants provided informed consent prior to the study and were compensated financially for their time. They reported being monolingual English speakers, who were currently residing in the UK, since the stimuli were normed for a UK population. Participants also reported having normal or corrected-to-normal vision. They had previously participated in online studies on the Prolific platform (<https://prolific.co/>) with an approval rating of at least 98%. The programme of research was reviewed by Cambridge Psychological Research Ethics Committee and received a favourable opinion (reference CPREC 2020.018). Procedures accorded with the Declaration of Helsinki.

Stimuli

The hermit words, invented neighbours and unfamiliar items were identical to those in Experiment 2. Pictures of typical and atypical referents from the same semantic category as the unfamiliar items were sourced from the internet. Typicality was determined in an independent ratings study with $n=27$ participants (see Supplementary Material). Following the approach used by Coutanche and Koch (2017), which relied on Ruts et al. (2004), we asked participants to rate the typicality of 162 exemplars across various semantic categories (e.g., bird, mammal, musical instrument) on a 20-point scale. From these ratings, we selected the 16 most typical and 16 most atypical items for each study category (natural and man-made). As in previous experiments, the assignment of stimulus set to study neighbour was counterbalanced across groups, with each set consisting of an equal number of 16 participants in each condition. The stimuli can be accessed here: <https://osf.io/vjcd6/>.

Procedure

The experimental procedure was identical to the FM group described in Experiment 2.

Analyses

As in previous analyses, items that were reported as known prior to the experiment were excluded from analyses of Recall, 3AFC, and the semantic decision task. The median

number of items reported as pre-experimentally known was 0, ranging up to 7.

The main focus of this experiment was the influence of referent typicality on lexical competition effects, specifically investigating whether the presence of atypical referents would enhance these competition effects. The data were first analysed by collapsing across the two within-participant's factors: semantic category of the study item and the semantic congruency with the hermit word. As in Coutanche and Thompson-Schill (2014), this analysis was conducted on mean RTs after trimming.

Subsequently, the analysis was expanded to include the within-participant factor of semantic congruency, to test the post hoc finding from Experiment 1 (though not found in pre-registered Experiment 2). Similar to previous experiments, we transformed the RTs instead of trimming them, using an inverse transform as employed in Experiment 1. Additionally, we examined the within-participant factor of semantic category of unknown object.

Finally, we combined the data from all FM groups across all three experiments to examine whether lexical competition effects were significant for the subset of items studied with atypical referents.

The R code and data are available in OSF: <https://osf.io/dpvbf/>.

Results

Main planned comparisons: competition RTs in implicit task

The bottom section of Table 1 presents the number and means of trimmed RTs for correct trials in the FM group, categorised based on whether a neighbour was studied. These values were consistent with those observed in previous experiments. These are also plotted in Panel J of Figure 2, along with the competition effect (studied – unstudied) in Panel K.

The primary analysis focused on the impact of referent typicality on lexical competition, averaged across stimulus type and semantic congruency. An ANOVA on trimmed RTs with factor Study and nested factor Referent Typicality (since by definition “unstudied” items do not have referents) revealed a significant interaction effect between Study and Referent Typicality, $F(1,126)=5.42$, $p=.02$, with longer RTs for neighbours of words studied with atypical than typical referents. However, there was no main effect of Study, $F(1,126)=0.79$, $p=.37$. A direct test of lexical competition effect for only the atypical referents failed to show significant slowing in RTs for studied items compared to unstudied items, $t(63)=0.44$, $p=.33$, one-tailed, with Cohen's d of 0.05. For typical referents, however, we found a significant speeding in RTs for studied items compared to unstudied items, $t(63)=-2.07$, $p=.043$, two-tailed, $d=0.26$. The latter might reflect a form of semantic priming, which is further explored below.

The secondary planned analysis aimed to replicate the congruency effect of semantic priming found in Experiment 1, but not Experiment 2, which predicts a speeding in RTs for studied items when the studied neighbour was congruent with the semantic judgement, but a slowing when incongruent. Though numerically in the expected direction, a one-tailed, paired t-test on inverse transformed RTs (collapsed over referent typicality and category) did not support the competition effect for Congruent trials ($M = -27.75$ ms) being significantly smaller than for Incongruent trials ($M = -16.17$ ms), $t(63) = 0.91$, $p = .18$, $d = -0.11$.

As in previous experiments, we also modelled the data using linear mixed effects models (see Supplementary Materials) and failed to find any significant effects. Analyses of the accuracy in the three explicit and implicit memory tasks have been reported in the Supplementary Materials.

Combined analysis across Experiments 1–3 of lexical competition in FM groups

In a final analysis, we combined the trial-averaged, trimmed RTs for the FM groups across Experiments 1–3 (combined $N = 152$). The combined effect size for the competition effect in the FM group was Cohen's $d = -0.06$, with a 95% confidence interval of $[-0.383, +0.258]$. The sample size of Experiment 1 of Coutanche and Thompson-Schill (2014) meant that it had 33% power to detect an effect size as small as $d = 0.438$, which is outside this confidence interval, and so according to Simonsohn (2015), our data are inconsistent with the notion that the true effect is large enough to have been detectable by the original Coutanche and Thompson-Schill (2014) experiment.

We also estimated Bayes Factors for T-tests as suggested by Rouder et al. (2009), using the default Cauchy prior scaled at $\sqrt{2}/2$ (medium scaling) (Morey & Rouder, 2015). The Bayes Factor represents the odds ratio for one hypothesis (e.g., the null hypothesis that there is no competition effect in the FM groups) versus another (e.g., the alternative, directional hypothesis that lexical competition is greater than 0). The Bayes Factor for the null was $BF_{01} = 18.51$, supporting no true effect.

Finally, for completeness, we also combined the congruency data with Natural items from the FM Groups across all three experiments. This showed a significant difference in the competition effect for congruent versus incongruent trials, $t(89) = -2.40$, $p = .009$, $BF_{10} = 3.44$, $d = -0.25$. However, this was not the case for Man-made items (across Experiments 2–3), $t(61) = 0.97$, $p = 0.83$, $BF_{01} = 13.3$, $d = -0.12$.

General discussion

There is a continuing debate about whether or not there are fast mapping (FM) processes that are distinct from

normal episodic encoding, particularly in adults (Cooper et al., 2019a). Coutanche and Thompson-Schill (2014) reported evidence for FM in healthy adults using an implicit measure of memory, namely reaction times (RTs) in a semantic decision task, which occurred after a fast mapping (FM) learning condition, but not after a more conventional explicit encoding (EE) condition. In this study, invented neighbours (e.g., “ganaxy”, “naskin”) of hermit words (“galaxy”, “napkin”) were learned as names of unfamiliar animals, with the logic that if these invented words are integrated into the lexicon, they will interfere with recognition of the original hermit words, i.e., slow-down RTs, as has been demonstrated in the classic “competition effect” on word recognition (Andrews, 1996; Bowers et al., 2005; Davis & Taft, 2005). Coutanche and Thompson-Schill (2014) reported such a competition effect in their FM condition, i.e., significant slowing of RTs to hermit words when a new neighbour had been incidentally learned as unfamiliar animal name during the study phase 10 min earlier. There was no such slowing in their EE condition, and the interaction was significant such that the competition effect was significantly greater in the FM than the EE condition. The authors concluded that only the FM condition enables same-day lexical integration.

However, in the present Experiments 1 and 2, powered at 82% and 98% respectively based on the effect size of $d = 0.69$ reported by Coutanche and Thompson-Schill (2014) for the interaction between competition effect and learning group, we were unable to replicate a greater competition effect in the FM than EE condition. Indeed, neither of these two experiments, nor Experiment 3 (which only included an FM condition), showed any significant RT evidence of same-day lexical integration following FM. Indeed, when combining the data across all three experiments, Bayes Factors supported the null hypotheses of no difference between hermit words with studied versus unstudied neighbours in the FM condition. This is despite the fact that the FM and EE tasks were sufficiently different for us to detect a difference in explicit memory performance, with the pattern in the current experiments echoing that in previous work (see Introduction), i.e., worse explicit memory in the FM task than EE task (with the FM-r task failing in between, as in Cooper et al., 2019b).

Though there was no evidence of Coutanche and Thompson-Schill's lexical competition effect, in a *post-hoc* analysis of our Experiment 1, we split the implicit test trials into natural versus man-made hermit words, and found an RT speeding for natural words (e.g., “galaxy”), but slowing for man-made words (e.g., “napkin”), in the FM condition, but not the EE condition. Since all to-be-learned items in Experiment 1 were animals, we speculated that this pattern reflects facilitatory versus inhibitory semantic priming effects for items from the same versus opposite category respectively (which cancel out to produce no net effect when averaged across word-type). Moreover, we suggested that the

presence of a semantically-related referent in the FM but not EE condition might explain why these priming effects were only found in the former (see Discussion to Experiment 1 for more details). We therefore expanded the design in Experiment 2 to test these hypotheses, by (1) adding a group for whom the unfamiliar objects were man-made rather than natural, and (2) including a third FM-r condition without a semantic referent. However, despite having 96% power to detect an effect size of Cohen's $d=0.55$ from Experiment 1, Experiment 2 failed to find any evidence in support of our post hoc semantic priming hypothesis, when collapsing over referent type, and failed to find any difference between whether or not a referent was included at all. Experiment 3 also found no evidence of this semantic congruency effect when collapsing over referent type.

However, it is worth noting that there was evidence of semantic congruency when restricting analysis of Experiments 2–3 to natural items (Table 2). This was the only referent category used in Experiment 1, and in Coutanche and Thompson-Schill (2014). Indeed, a combined analysis across all three of the present experiments provided strong evidence of semantic congruency effects for natural items. This was not the case for man-made items, which explains why the effect of congruency did not reach significance when averaging across these two referent types. We do not know why the category of the items makes such a difference to congruency effects – it might be because participants re-interpret the semantic decision task from “natural or man-made?” to “is it natural: yes/no”, with congruency only affecting affirmative responses – but this a topic for future experiments, since the focus here was on the basic lexical competition effect collapsed across congruency.

Another plausible explanation for any differences between our Experiments 2 and 3 in comparison to our Experiment 1, as well as our Experiments 2 and 3 and those conducted by Coutanche and Thompson-Schill (2014), could be attributed to less reliable data are collected online (rather than in the laboratory). For example, there are concerns over measurement error in recording RTs, given the range of hardware/browsers used by participants. However, these concerns do not seem to apply to within-participant measures, which subtract out systematic differences in RTs across participants (see Anwyll-Irvine et al., 2021; Bridges et al., 2020), which may be why online studies have replicated many RT differences detected in laboratory settings (e.g., Kochari, 2019; Semmelmann & Weigelt, 2017). Thus we think an RT measurement error is unlikely to be a problem. It is also possible that participants were not as motivated when tested online, or took extra breaks between sessions, which would be consistent for example with numerically worse accuracy for the semantic decision task in Experiments 2 and 3 than Experiment 1. However, RTs on that task, as well as accuracy on the explicit memory test, were comparable across our experiments, and comparable

with those in Coutanche and Thompson-Schill (2014). Thus we do not think the online nature of data collection in Experiments 2 and 3 is a likely problem.

It is possible that minor methodological details caused our failure to replicate Coutanche and Thompson-Schill's (2014) findings. For example, the presence of explicit memory tests (recall and 3AFC) before the implicit test could have affected performance (RTs) on the implicit test, somehow eliminating the competition effect. While this might explain the lack of a competition effect in our FM task, this does not explain the presence of a competition effect in Experiment 1 of Coutanche and Thompson-Schill (2014), which used the same test order as us. However, future studies might be needed to systematically investigate and account for potential carry-over effects that may arise from the test order. Another possibility is the semantic trait of the participants. As mentioned in the Results section of Experiment 1, Coutanche and Koch (2017) reported that the competition effect was stronger in participants with high “semantic trait score”.⁹ We do not have such trait scores for the present participants, but we see no reason why our participants should have lower (or higher) such scores than the participants in the original Coutanche and Thompson-Schill (2014) study. While participants in the FM group of Experiment 1 happened to have lower verbal intelligence scores (on WAIS), linear models provided no evidence that verbal IQ modulated the size of any lexical competition effects.

Finally, a third possible methodological difference was the choice of referent object during study. Coutanche and Koch (2017) claimed that the less typical the known referent in the FM condition, the more evidence there should be for same-day lexicalisation (see also Coutanche, 2019). Experiments 1 and 2 used our own set of normed stimuli from Greve et al. (2014) and Cooper et al. (2019b), to ensure that unknown and known items fulfilled their brief for UK rather than US participants. Post-hoc analyses showed that the typicality of our known referents in Experiments 1 and 2 was significantly lower than the typicality of the referents in Coutanche and Thompson-Schill (2014), suggesting that we should have seen larger lexical competition effects. We then also conducted Experiment 3 to specifically investigate the effects of referent typicality on lexical competition in the FM group, using more extreme atypical and typical referents, like in Coutanche and Koch (2017). However, the results of Experiment 3 failed to provide any evidence that more atypical referents result in larger lexical competition effects. While we saw significant speeding in RTs in the typical condition, resulting in a significant difference between the typical and atypical conditions, we did not observe significant lexical competition in the atypical condition. The facilitation effect in the typical condition was also documented in a subset of the participants in Coutanche and Koch (2017), but this speeding is not the effect expected from Coutanche and Thompson-Schill's theory (2014; Coutanche & Koch, 2017) and provides no

evidence of fast mapping. Moreover, the *a priori* reason for this boundary condition of typicality is debatable: some theories of semantic activation would seem to predict that a more, not less, typical item should support lexical integration (Rosch & Mervis, 1975; see also, Cooper et al., 2019c; and Mak, 2019; though see Coutanche, 2019; and Zaiser et al., 2019, for counter-arguments). Thus, we remain sceptical that differences in semantic traits of participants or typicality of the referent object caused the failure to replicate Coutanche and Thompson-Schill (2014).

Despite our difficulties replicating the effect, it is of course still possible that an effect exists, but its effect size is likely to be smaller than the medium-to-large effect originally reported by Coutanche and Thompson-Schill (2014). There is a known replication crisis in psychology due to a number of factors (Wiggins & Christopherson, 2019), one of which is that published effect sizes are likely to be larger than the true effect size (e.g., owing to a bias to only report significant results; Anderson & Maxwell, 2017). If the FM task (but not EE task) does cause same-day lexicalisation, but the effect size is much smaller, it becomes less interesting in terms of practical implications (Diener & Biswas-Diener, 2018), particularly if it is sensitive to boundary conditions such as participant traits (see previous paragraph). Nonetheless, a small but non-zero effect size still has theoretical implications (e.g., over whether the FM condition engenders a qualitatively different type of memory encoding; see Introduction). Thus while we (or a future meta-analysis) could calculate an updated effect size across multiple experiments, we prefer the “categorical” approach of Simonsohn (2015), or Bayes Factor approach, i.e., to claim that it is more likely that there is no effect (i.e., true effect size is 0).

The present experiments only question the behavioural evidence for FM as a distinct type of learning. Neuroimaging experiments have suggested distinct neural components. Scanning brain activity with fMRI during the study phase could reveal differences simply due to the different stimulus arrays (e.g., presence of a referent in the FM but not EE condition), while differences in the test phase could reflect incidental retrieval of the referent, even if that retrieval does not affect memory for the target name (see Cooper et al., 2019a). Zaiser et al. (2022) reported greater fMRI response in perirhinal cortex associated with later remembered versus forgotten names for items with high compared to low feature overlap, but did not include an EE condition. Two event-related potentials (ERP) studies (Shtyrov et al., 2021, 2022) reported changes in the ERP to spoken words after, versus before, being studied under FM or EE conditions. Though these studies did not test lexical competition effects via RTs like in Coutanche and Thompson-Schill (2014), the authors claimed that different study conditions caused different changes in the ERP to the newly learned word.

Indeed, in the later study (Shtyrov et al., 2022), the magnitude of the ERP change in an early component predicted explicit memory performance in the FM task, whereas that

of a later component predicted explicit memory in the EE task.¹⁰ While these ERP studies do suggest that different types of representations are established following FM or EE learning conditions, they do not demonstrate that these differences reflect fast cortical mapping, e.g., lexicalisation necessary to produce the RT competition effect, and future work is needed to check the ERP differences do not reflect differences in memory strength (e.g., stronger in EE condition, given that explicit memory was better for EE than FM, as in all previous studies in this paradigm), or even incidental retrieval of different source (context) information from the different study conditions.

Finally, we should clarify that we are not claiming that fast cortical mapping (FCM) does not exist. Indeed, we believe that some types of FCM does occur, i.e., rapid but long-lasting changes in the cortex that are independent of the hippocampus (e.g., in long-term priming). We also accept that one could investigate a hypothetical process of “fast mapping” without committing to any specific neural bases (e.g., without the “cortical” vs “hippocampal” distinction), as is common in the developmental literature (see Cooper et al., 2019c, for further discussion). Rather, what we are claiming is that there is little evidence for such a distinct process provided by the specific FM paradigm introduced by Sharon et al. (2011) and extended to implicit measures by Coutanche and Thompson-Schill (2014).

In conclusion, we failed to replicate any difference between FM and EE conditions using an implicit measure of same-day lexicalisation; indeed, we failed to find any evidence of same-day lexicalisation in any condition. This further questions the existence of a fast mapping process uniquely revealed by the current FM paradigm, at least in adults.

Notes

1. When one condition (FM) produces lower overall performance than another (EE), one must be careful that smaller or absent effects of another variable (like sleep or retention interval) on the former (FM) do not simply reflect reduced range for finding an effect (e.g., “attenuation” or “floor” effect).
2. A model with random slopes too did not converge.
3. Thirteen participants were removed either because performance did not meet data quality checks specified in our pre-registration document, e.g., low study phase performance, self-report that a participant did not understand some instructions, or erroneous responses in the free-type boxes, e.g., the recall test. Replacements were: Five participants in FM-Natural, four in the FM-r-Natural, two in FM-Man-made, one in FM-r-Man-made, and one in EE-Man-made.
4. The pre-registration incorrectly stated $N = 30$ per group, rather than $N = 60$, forgetting that this test can in principle collapse over the referent category.
5. This parametrisation differed from what was pre-registered, where we stated that the factors would refer to semantic category of referent and semantic category of hermit word, but the equivalent, re-parametrisation by congruency aids comparison with Experiment 1.

6. Results were similar even when both sets of stimuli were rated by the same group of participants (see Supplementary Materials)
7. 2 participants faced technical problems, and 4 participants reported pre-experimental knowledge of the “unknown” images.
8. Age and sex were unavailable for 1 participant.
9. Though they do not report an analysis that collapses over participant traits and all conditions, thus not reporting a complete replication of the lexical competition effect under FM as originally reported by Coutanche and Thompson-Schill (2014).
10. It is worth noting the different ERP correlates came from separate regressions within each condition, with no direct comparison of regression slopes across conditions; statistical evidence for such an interaction would bolster claims for a neural dissociation.

Open Scholarship

This article has earned the Center for Open Science badges for Open Data, Open Materials and Preregistered. The data and materials are openly accessible at <https://osf.io/dpvbf/>, <https://osf.io/vjcd6/> and <https://osf.io/atkp4>, <https://osf.io/p7s4f>, <https://osf.io/wqf83>.

Acknowledgements

We are very grateful to our volunteers for enabling this research and to Dr J.S.H. Taylor for sharing resources, specifically, the stimuli set of novel object pictures used in Experiment 2. We thank Josefina Weinerova and Jordan Punshon-Smith for their help with data collection and collation, and an internal reviewer, Dr Dennis Norris, for comments and suggestions on an earlier version of this report. We are also grateful to three anonymous reviewers on a previous version of this manuscript. For the purpose of open access, the author has applied a Creative Commons Attribution (CC BY) licence to any Author Accepted Manuscript version arising from this submission.

Data availability statement

Stimuli are available on the Open Science Framework (OSF): <https://osf.io/vjcd6/>. Data and the R code for analysis are available on OSF: <https://osf.io/dpvbf/>. Prior to starting data collection or analyses, Experiment 2 was pre-registered on OSF using a replication registration report template: the pre-registration for conditions FM and EE is located at <https://osf.io/atkp4>; the pre-registration for the FM-r condition is located at <https://osf.io/p7s4f>. Experiment 3 was pre-registered on OSF at <https://osf.io/wqf83>.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This work was supported by UK Medical Research Council (MRC) [grant number SUAG/046/G101400]. K.G. is supported by Hezkuntza Saila, Eusko Jaurlaritza [grant number POS_2022_2_0025]. R.T. is supported by the British Academy [grant number SUA1/028 RG94188].

References

Anderson, S. F., & Maxwell, S. E. (2017). Addressing the “replication crisis”: Using original studies to design replication studies with

- appropriate statistical power. *Multivariate Behavioral Research*, 52(3), 305–324. <https://doi.org/10.1080/00273171.2017.1289361>
- Andrews, S. (1996). Lexical retrieval and selection processes: Effects of transposed-letter confusability. *Journal of Memory and Language*, 35(6), 775–800. <https://doi.org/10.1006/jmla.1996.0040>
- Anwyl-Irvine, A., Dalmaijer, E. S., Hodges, N., & Evershed, J. K. (2021). Realistic precision and accuracy of online experiment platforms, web browsers, and devices. *Behavior Research Methods*, 53(4), 1407–1425. <https://doi.org/10.3758/s13428-020-01501-5>
- Baker, D. H., Vilidaite, G., Lygo, F. A., Smith, A. K., Flack, T. R., Gouws, A. D., & Andrews, T. J. (2021). Power contours: Optimising sample size and precision in experimental psychology and human neuroscience. *Psychological Methods*, 26(3), 295–314. <https://doi.org/10.1037/met0000337>
- Bowers, J. S., Davis, C. J., & Hanley, D. A. (2005). Interfering neighbours: The impact of novel word learning on the identification of visually similar words. *Cognition*, 97(3), 45–54. <https://doi.org/10.1016/j.cognition.2005.02.002>
- Bridges, D., Pitiot, A., MacAskill, M. R., & Peirce, J. W. (2020). The timing mega-study: Comparing a range of experiment generators, both lab-based and online. *PeerJ*, 8, e9414. <https://doi.org/10.7717/peerj.9414>
- Carey, S. (1978). The child as word learner. In M. Halle, J. Bresnan, & G. A. Miller (Eds.), *Linguistic theory and psychological reality* (pp. 264–293). MIT Press.
- Carey, S. (2010). Beyond fast mapping. *Language Learning and Development*, 6(3), 184–205. <https://doi.org/10.1080/15475441.2010.484379>
- Carey, S., & Bartlett, E. (1978). Acquiring a single new word. *Child Language Development*, 15, 17–29.
- Cattell, R. B., & Cattell, A. K. S. (1960). *Cattell test of “g”: Culture fair, scale 2, form A*. Institute for Personality and Ability Testing, Inc.
- Collins, A. M., & Loftus, E. F. (1975). A spreading-activation theory of semantic processing. *Psychological Review*, 82(6), 407–428. <https://doi.org/10.1037/0033-295X.82.6.407>
- Cooper, E., Greve, A., & Henson, R. N. (2019a). Little evidence for fast mapping (FM) in adults: A review and discussion. *Cognitive Neuroscience*, 10(4), 196–209. <https://doi.org/10.1080/17588928.2018.1542376>
- Cooper, E., Greve, A., & Henson, R. N. (2019b). Investigating fast mapping task components: No evidence for the role of semantic referent nor semantic inference in healthy adults. *Frontiers in Psychology*, 10(394). <https://doi.org/10.3389/fpsyg.2019.00394>
- Cooper, E., Greve, A., & Henson, R. N. (2019c). Response to commentaries on our review of fast mapping in adults. *Cognitive Neuroscience*, 10(4), 237–240. <https://doi.org/10.1080/17588928.2019.1651709>
- Coutanche, M. N. (2019). Addressing misconceptions of fast mapping in adults. *Cognitive Neuroscience*, 10(4), 226–228. <https://doi.org/10.1080/17588928.2019.1593955>
- Coutanche, M. N., & Koch, G. E. (2017). Variation across individuals and items determine learning outcomes from fast mapping. *Neuropsychologia*, 106, 187–193. <https://doi.org/10.1016/j.neuropsychologia.2017.09.029>
- Coutanche, M. N., & Thompson-Schill, S. L. (2014). Fast mapping rapidly integrates information into existing memory networks. *Journal of Experimental Psychology: General*, 143(6), 2296–2303. <https://doi.org/10.1037/xge0000020>
- Coutanche, M. N., & Thompson-Schill, S. L. (2015). Rapid consolidation of new knowledge in adulthood via fast mapping. *Trends in Cognitive Sciences*, 19(9), 486–488. <https://doi.org/10.1016/j.tics.2015.06.001>
- Davis, C. J., & Taft, M. (2005). More words in the neighborhood: Interference in lexical decision due to deletion neighbors. *Psychonomic Bulletin & Review*, 12(5), 904–910. <https://doi.org/10.3758/BF03196784>
- de Leeuw, J. R. (2015). jsPsych: A JavaScript library for creating behavioral experiments in a web browser. *Behavior Research Methods*, 47(1), 1–12. <https://doi.org/10.3758/s13428-014-0458-y>

- Diener, E., & Biswas-Diener, R. (2018). The replication crisis in psychology. In R. Biswas-Diener & E. Diener (Eds.), *Noba textbook series: Psychology* (pp. 6–18). <https://nobaproject.com/modules/the-replication-crisis-in-psychology>
- Elward, R. L., Dziezic, A. M., & Vargha-Khadem, F. (2019). Little evidence for fast mapping in adults with developmental amnesia. *Cognitive Neuroscience*, 10(4), 215–217. <https://doi.org/10.1080/17588928.2019.1593123>
- Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, 39(2), 175–191. <https://doi.org/10.3758/BF03193146>
- Goshen-Gottstein, Y., Moscovitch, M., & Melo, B. (2000). Intact implicit memory for newly formed verbal associations in amnesic patients following single study trials. *Neuropsychology*, 14(4), 570–578. <https://doi.org/10.1037//0894-4105.14.4.570>
- Greve, A., Cooper, E., & Henson, R. N. (2014). No evidence that ‘fast-mapping’ benefits novel learning in healthy older adults. *Neuropsychologia*, 60(1). <https://doi.org/10.1016/j.neuropsychologia.2014.05.011>
- Himmer, L., Müller, E., Gais, S., & Schönauer, M. (2017). Sleep-mediated memory consolidation depends on the level of integration at encoding. *Neurobiology of Learning and Memory*, 137, 101–106. <https://doi.org/10.1016/j.nlm.2016.11.019>
- Kapnoula, E. C., Packard, S., Gupta, P., & McMurray, B. (2015). Immediate lexical integration of novel word forms. *Cognition*, 134, 85–99. <https://doi.org/10.1016/j.cognition.2014.09.007>
- Kochari, A. R. (2019). Conducting web-based experiments for numerical cognition research. *Journal of Cognition*, 2(1), 39. <https://doi.org/10.5334/joc.85>
- Lange, K., Kühn, S., & Filevich, E. (2015). “Just Another Tool for Online Studies” (JATOS): An easy solution for setup and management of web servers supporting online studies. *PLoS One*, 10(6), e0130834. <https://doi.org/10.1371/journal.pone.0130834>
- Li, C., Hu, Z., & Yang, J. (202). Rapid acquisition through fast mapping: Stable memory over time and role of prior knowledge. *Learning & Memory*, 27(5), 177–189. <https://doi.org/10.1101/lm.050138.119>
- Lindsay, S., & Gaskell, M. G. (2013). Lexical integration of novel words without sleep. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 39(2), 608–622. <https://doi.org/10.1037/a0029243>
- Mak, M. H. C. (2019). Why and how the co-occurring familiar object matters in fast mapping (FM)? Insights from computational models. *Cognitive Neuroscience*, 10(4), 229–231. <https://doi.org/10.1080/17588928.2019.1593121>
- McGregor, K. K., Eden, N., Arbis-Kelm, T., & Oleson, J. (2020). The fast-mapping abilities of adults with developmental language disorder. *Journal of Speech, Language, and Hearing Research*, 63(9), 3117–3129. https://doi.org/10.1044/2020_JSLHR-19-00418
- McKenzie, S., & Eichenbaum, H. (2011). Neuron perspective consolidation and reconsolidation: Two lives of memories? *Neuron*, 71(2), 224–233. <https://doi.org/10.1016/j.neuron.2011.06.037>
- McNamara, T. P. (2005). What is semantic priming and why should anyone care about it? In T. P. McNamara (Ed.), *Semantic priming: Perspectives from memory and word recognition. (Essays in cognitive psychology)* (pp. 3–8). Psychology Press.
- Merhav, M., Karni, A., & Gilboa, A. (2014). Neocortical catastrophic interference in healthy and amnesic adults: A paradoxical matter of time. *Hippocampus*, 24(12), 1653–1662. <https://doi.org/10.1002/hipo.22353>
- Morey, R. D., & Rouder, J. N. (2015). *Computation of Bayes Factors for common designs*. Package ‘BayesFactor’ for R. <https://cran.r-project.org/web/packages/BayesFactor/BayesFactor.pdf>.
- Norman, K. A., & O’Reilly, R. C. (2003). Modeling hippocampal and neocortical contributions to recognition memory: A complementary-learning-systems approach. *Psychological Review*, 110(4), 611–646. <https://doi.org/10.1037/0033-295X.110.4.611>
- O’Connor, R. J., Lindsay, S., Mather, E., & Riggs, K. J. (2019). Why would a special FM process exist in adults, when it does not appear to exist in children? *Cognitive Neuroscience*, 10(4), 221–222. <https://doi.org/10.1080/17588928.2019.1574260>
- O’Connor, R. J., & Riggs, K. J. (2019). Adult fast-mapping memory research is based on a misinterpretation of developmental-word-learning data. *Current Directions in Psychological Science*, 28(6), 528–533. <https://doi.org/10.1177/0963721419858426>
- Psychological Software Tools, Inc. (2012). *E-Prime 2.0*. Pittsburg, PA.
- Ratcliff, R. (1993). Methods for dealing with reaction time outliers. *Psychological Bulletin*, 114(3), 510–532. <https://doi.org/10.1037/0033-2909.114.3.510>
- R Core Team. (2020). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. <https://www.R-project.org/>
- Rosch, E., & Mervis, C. B. (1975). Family resemblances: Studies in the internal structure of categories. *Cognitive Psychology*, 7(4), 573–605. [https://doi.org/10.1016/0010-0285\(75\)90024-9](https://doi.org/10.1016/0010-0285(75)90024-9)
- Rouder, J. N., Speckman, P. L., Sun, D., Morey, R. D., & Iverson, G. (2009). Bayesian t tests for accepting and rejecting the null hypothesis. *Psychonomic Bulletin & Review*, 16(2), 225–237. <https://doi.org/10.3758/PBR.16.2.225>
- RStudio Team. (2020). *Rstudio: Integrated development for R*. RStudio, PBC. <http://www.rstudio.com/>
- Ruts, W., De Deyne, S., Ameel, E., Vanpaemel, W., Verbeemen, T., & Storms, G. (2004). Dutch norm data for 13 semantic categories and 338 exemplars. *Behavior Research Methods, Instruments, & Computers*, 36(3), 506–515. <https://doi.org/10.3758/BF03195597>
- Schacter, D. L., & Tulving, E. (1994). In D. L. Schacter, & E. Tulving (Eds.), *Memory systems* (pp. 1–38). MIT Press.
- Semmelmann, K., & Weigelt, S. (2017). Online psychophysics: Reaction time effects in cognitive experiments. *Behavior Research Methods*, 49(4), 1241–1260. <https://doi.org/10.3758/s13428-016-0783-4>
- Sharon, T., Moscovitch, M., & Gilboa, A. (2011). Rapid neocortical acquisition of long-term arbitrary associations independent of the hippocampus. *Proceedings of the National Academy of Sciences*, 108(3), 1146–1151. <https://doi.org/10.1073/pnas.1005238108>
- Shtyrov, Y., Filippova, M., Blagovechtchenski, E., Kirsanov, A., Nikiforova, E., & Shcherbakova, O. (2021). Electrophysiological evidence of dissociation between explicit encoding and fast mapping of novel spoken words. *Frontiers in Psychology*, 12, 143. <https://doi.org/10.3389/fpsyg.2021.571673>
- Shtyrov, Y., Filippova, M., Perikova, E., Kirsanov, A., Shcherbakova, O., Blagovechtchenski, E. (2022). Explicit encoding vs. fast mapping of novel spoken words: Electrophysiological and behavioural evidence of diverging mechanisms. *Neuropsychologia*. <https://doi.org/10.1016/j.neuropsychologia.2022.108268>.
- Simonsohn, U. (2015). Small telescopes: Detectability and the evaluation of replication results. *Psychological science*, 26(5), 559–569. <https://doi.org/10.1177/0956797614567341>.
- Smith, C. N., Urgolites, Z. J., Hopkins, R. O., & Squire, L. R. (2014). Comparison of explicit and incidental learning strategies in memory-impaired patients. *Proceedings of the National Academy of Sciences*, 111(1), 475–479. <https://doi.org/10.1073/pnas.1322263111>
- Squire, L. R., & Bayley, P. J. (2007). The neuroscience of remote memory. *Current Opinion in Neurobiology*, 17(2), 185–196. <https://doi.org/10.1016/j.conb.2007.02.006>
- Taylor, J. S. H., Rastle, K., Davis, M.H. (2014). Distinct neural specializations for learning to read words and name objects. *Journal of Cognitive Neuroscience*, 26(9), 2128–2154. https://doi.org/10.1162/jocn_a_00614
- Tulving, E., & Schacter, D. L. (1990). Priming and human memory systems. *Science*, 19(247), 301–306. <https://doi.org/10.1126/science.2296719>
- Warren, D. E., & Duff, M. C. (2014). Not so fast: Hippocampal amnesia slows word learning despite successful fast mapping. *Hippocampus*, 24(8), 920–933. <https://doi.org/10.1002/hipo.22279>

- Warren, D. E., Tranel, D., & Duff, M. C. (2016). Impaired acquisition of new words after left temporal lobectomy despite normal fast-mapping behavior. *Neuropsychologia*, *80*, 165–175. <https://doi.org/10.1016/j.neuropsychologia.2015.11.016>
- Wechsler, D. (1999). *Wechsler abbreviated scale of intelligence*. Pearson.
- Wiggins, B. J., & Christopherson, C. D. (2019). The replication crisis in psychology: An overview for theoretical and philosophical psychology. *Journal of Theoretical and Philosophical Psychology*, *39*(4), 202–217. <https://doi.org/10.1037/teo0000137>
- Zaiser, A-K, Bader, R., & Meyer, P. (2022). High feature overlap reveals the importance of anterior and medial temporal lobe structures for learning by means of fast mapping. *Cortex*, *146*, 74–88. <https://doi.org/10.1016/j.cortex.2021.07.017>
- Zaiser, A-K, Meyer, P., & Bader, R. (2019). Evidence for fast mapping in adults – Moderating factors yet need to be identified. *Cognitive Neuroscience*, *10*(4), 232–233. <https://doi.org/10.1080/17588928.2019.1605986>
- Zaiser, A-K, Meyer, P., & Bader, R. (2021). High feature overlap and incidental encoding drive rapid semantic integration in the fast mapping paradigm. *Journal of Experimental Psychology: General*. Advance online publication. <https://doi.org/10.1037/xge0001070>