Memory & Cognition
https://doi.org/10.3758/s13421-025-01823-9

Pretesting effects for facts reflect lexical over semantic or structural features

Steven C. Pan[1] · **Faith Amanda Siauw**[1]

Received: 21 July 2025 / Accepted: 10 November 2025

Abstract
To what extent does *pretesting* – the learning technique of guessing to-be-learned content before studying the correct answers – enhance factual knowledge, and does its effectiveness depend on which part of a fact is pretested? We investigated memory for different parts of facts when pretesting involved either the subject (e.g., "_____ *is the goddess of agriculture*") or a key term from the predicative (e.g., "*Demeter is the goddess of* _____"), compared to studying intact facts. Experiments 1a–4 revealed asymmetrical learning patterns: Pretesting on the subject enhanced memory for the predicative key term, but not consistently for the subject itself, whereas pretesting on the predicative key term, at best, enhanced memory for that key term but not for the subject. When word frequency among subject and predicative terms were closely matched in Experiments 5–7, however, pretesting yielded highly symmetrical learning – enhancing memory for both terms regardless of which one was pretested. That enhancement occurred despite differences in associative strength and semantic structure, whereas syntactic and other structural influences, including extra cue words, were ruled out. Together, these findings reveal that pretesting can improve memory for both tested and untested fact terms, with lexical properties exerting strong influences on the scope and magnitude of improvement. The effects of pretesting on facts therefore depend on how easily learners can access and encode fact elements, and accordingly, which terms to pretest should be considered carefully when using pretesting for fact learning.

Keywords Pretesting effect · Transfer · Cued recall · Test-enhanced learning

Introduction

A growing body of research reveals that making guesses about to-be-learned information before studying the correct answers, or *pretesting*, can enhance memory (Kornell et al., 2009; Richland et al., 2009; for reviews, see Kornell & Vaughn, 2016; Mera et al., 2021; Pan & Carpenter, 2023). In a typical pretesting experiment, participants guess the answers to a series of pretest questions about unfamiliar topics (e.g., "_____ *is the capital of Suriname*"), often doing so incorrectly, and receive immediate correct answer feedback (e.g., "*Paramaribo*"). This pretesting condition is compared to a studying (or reading) condition, where participants study or read correct information without guessing or feedback. On a subsequent criterial test, the pretesting condition often exhibits better recall than the studying condition – a phenomenon known as the *pretesting effect.*

The pretesting effect has been successfully demonstrated with stimuli such as paired associate words (e.g., Hays et al., 2013), trivia facts (e.g., Kornell et al., 2009), and text passages (e.g., Richland et al., 2009), and in most cases, on criterial tests assessing memory for directly pretested information (i.e., the answers to pretest questions – e.g., "*Paramaribo*"). Its occurrence, however, is not always consistent, as illustrated by cases where pretesting has failed to enhance learning (e.g., with live lessons as in Geller et al., 2017). Moreover, the extent to which the effect transfers to untested materials is even less clear. For example, if one is pretested on "_____ *is the capital of Suriname,*" and then studies the correct answer, does doing so enhance memory for surrounding information – such as "*capital*" or "*Suriname*"? That question has yet to be directly addressed, with few relevant studies and mixed findings to date.

✉ Steven C. Pan
scp@nus.edu.sg

[1] Department of Psychology, Faculty of Arts and Social Sciences, National University of Singapore, 9 Arts Link, Singapore City 117572, Singapore

Pretesting effects for tested and untested materials

In two studies with simple verbal stimuli, the memorial benefits of pretesting have transferred successfully to untested materials. Hays et al. (2013) had participants engage in pretesting on the target word of paired associates (e.g., *frog–???*), then attempt to recall the untested cue word on a criterial test (e.g., *???–pond*). When immediate correct answer feedback was provided, as is common for pretesting with such materials, memory for cue and target words was better after pretesting than after studying intact pairs. Similarly, Pan et al. (2019) had participants engage in pretesting on one word per word triplet (e.g., *gift, rose,???*), then take a criterial test in which each word from each triplet was assessed separately (e.g., *???, rose, wine* on one trial; *gift,???, wine* on another, etc.). Pretesting improved memory for both tested and untested words relative to studying, providing further evidence of transfer.

In contrast, with text passages – where participants attempt pretest questions prior to reading a passage, in which the correct answers are discoverable – pretesting often improves memory for directly tested information only. For instance, in a series of experiments by Richland et al., (2009; see also James & Storm, 2019), pretesting on a text passage about achromatopsia enhanced memory for tested material but yielded no advantage (or disadvantage) for untested portions compared to studying. Although some exceptions have been reported (e.g., Pan & Sana, 2021) and the lack of correct answer feedback may be a contributing factor, a recent meta-analysis concluded that the pretesting effect for text passages is statistically significant only for tested information (St. Hilaire et al., 2024).

Together, the foregoing studies suggest that the capacity of pretesting to enhance recall of untested but nonetheless presented content likely varies according to the materials being learned (e.g., Hays et al., 2013; Pan et al., 2019; Richland et al., 2009). Given the limited evidence to date, however, further research is needed to clarify the conditions under which such transfer might occur. Hence, the present study investigated pretesting effects for tested and untested materials using facts, which are stimuli of intermediate complexity between paired associates or word triplets and text passages, with distinct structural, semantic, and lexical properties.

Relevant theorizing from the pretesting and other literatures

Prominent theories of the pretesting effect, largely developed to explain findings involving paired associate learning, differ in both explanatory mechanisms and predictions about the fate of tested and untested materials. Most theories are only explicit about how tested materials are affected. For instance, mediator-based accounts (Bartl et al., 2024; Carpenter, 2011; Kornell et al., 2009) propose that pretesting prompts the generation of mediator words that mentally link cues to targets, with the mediators aiding recall of targets (e.g., given *doctor-???,* for which the answer is *nurse*, one might mentally generate *patient*). An interpretation of that account is that these mediators might also facilitate recall of cues, although that possibility was not originally specified. Search set accounts (e.g., Grimaldi & Karpicke, 2012), meanwhile, posit that pretesting in the context of cue-target learning can activate a set of semantically related concepts – perhaps through spreading activation within semantic memory (Collins & Loftus, 1975) – and when feedback is provided, processing of the correct cue-target pair is enhanced (e.g., given *doctor-???*, one might mentally activate such concepts as *patient, hospital, nurse*). This process applies only when pretesting occurs on semantically related cues and targets. One interpretation of this account is that pretesting may strengthen memory for cues and their corresponding targets, although it does not directly make that claim.

Another influential theory, the memory-for-change account (Wahlheim & Jacoby, 2013; see also Jacoby & Wahlheim, 2013; Mera et al., 2021; Wahlheim et al., 2019), which drew inspiration from Hintzman's (2011) recursive reminding hypothesis, implies that the experience of generating an error during a pretest and then learning the correct answer becomes integrated within the same episodic memory trace, with that trace supporting improved recall of directly pretested material (e.g., forming a memory of guessing the word *medicine* in response to *doctor-???*, and then learning that the correct answer is *nurse*). If cue information is encoded in this integrated event, then transfer to untested material might be possible. Finally, attentional accounts involving paired associate learning propose that pretesting enhances attention to targets presented as feedback (e.g., greater attention to the word *nurse* when it is shown as feedback), with memory for targets benefiting in a selective manner (Potts & Shanks, 2014; Seabrooke et al., 2021); accounts involving text passage learning suggest that pretesting guides attention to relevant passage content, potentially benefiting untested material as well (e.g., Sana & Carpenter, 2023; see also Little & Bjork, 2016). In the latter case, pretesting might prompt greater attention to cue information from facts than studying alone, and if so, memory for untested cue information could be enhanced.

Relatedly, in the retrieval practice (i.e., recall testing) literature, it has been observed that performing cued recall of previously studied facts (e.g., Hinze & Wiley, 2011) and word triplets (e.g., Pan et al., 2016a, 2016b) typically enhances memory for tested words and not untested words

or cues, relative to a restudy condition. That specificity of learning reflects a recurring theme in memory research – as exemplified by the classic theory of identical elements (Thorndike & Woodworth, 1901), the encoding specificity principle (Tulving & Thomson, 1973), and identical elements models for multi-element stimuli (Rickard et al., 1994) – which emphasize that differences in contextual elements or cue-target combinations between learning and testing reduce transfer. An exception involves retrieval practice on word pairs, which can enhance memory for both tested and untested words within the same pair (e.g., Carpenter et al., 2006), and there are cases where retrieval practice can support more distant forms of transfer (e.g., classifying novel exemplars, solving application and inference questions; Butler, 2010; Jacoby et al., 2010; for review, see Pan & Rickard, 2018). If the findings with retrieval practice on facts and word triplets are applicable to pretesting, however, then one might expect memorial benefits for directly tested content only.

A potentially important distinction between pretesting and other test-based techniques such as retrieval practice is that pretesting may rely more heavily on semantic memory than episodic memory, at least during the initial pretesting event (Huelser & Metcalfe, 2012; Knight et al., 2012; Pan et al., 2019; cf. Metcalfe & Huelser, 2020). Pretesting lacks an initial study episode, and because target information has not yet been studied, learners must instead draw on general knowledge and the cues in a given pretest question to generate guesses. Accordingly, pretesting might engage different memory dynamics that are not subject to hippocampal-dependent pattern separation processes (Schlichting et al., 2015) that can selectively enhance targets over cues in episodic memory, yielding more transferrable learning. Given these possibilities and aforementioned theoretical accounts, further research on pretesting effects for tested and untested information stands to yield insights.

Impacts of stimulus complexity and semantic, structural, or lexical features

Stimulus complexity can influence the conditions under which pretesting enhances learning, including the optimal timing of correct answer feedback. For simple materials such as word pairs, immediate feedback is essential (Grimaldi & Karpicke, 2012; Hays et al., 2013; Kornell et al., 2009; Vaughn & Rawson, 2012; cf. Zawadzka et al., 2023), whereas for more complex materials such as facts, feedback can be delayed by several minutes or even 24 h (Kornell, 2014). This difference may stem from richer semantic networks activated during fact learning (Kornell, 2014), which might foster deeper processing and support greater transfer to untested materials. As previously noted, however, evidence to date involving word pairs and triplets versus text passages is more evocative of the reverse pattern, suggesting that factors other than stimulus complexity might influence pretesting effects for untested materials more strongly.

Syntactic and other structural factors, including grammatical role, might be such factors. Most facts include a *subject* – the entity the fact is about (e.g., "*Paramaribo*") – which typically serves as the focus of the fact. The subject is usually paired with a *predicate*, which often contains a *predicative* phrase (e.g., "*capital of Suriname*") that renames or describes the subject. Together, the subject and predicative represent crucial information, but their different grammatical roles may affect how they are learned. Other influential structural factors might include the presence of modifier words, which in other experimental contexts has been known to act as retrieval cues, affecting encoding or recall of associated nouns and verbs (e.g., Kasahara & Yanagisawa, 2024; Lockhart, 1969). Syntactic structure, including the linear arrangement of terms, may be impactful as well (Weyerts et al., 2002); for instance, in sentence recall, words at the beginning and end tend to be remembered better than those in the middle (Wearing, 1971).

Semantic characteristics such as associative strength between terms may also be influential. It is well established that pretesting fails to enhance subsequent cued recall of targets from paired associates when the cue and target are semantically unrelated (Grimaldi & Karpicke, 2012; Huelser & Metcalfe, 2012; Knight et al., 2012; see also Potts & Shanks, 2014; Seabrooke et al., 2021). Facts typically contain semantically associated terms, but these associations can be asymmetrical, with the forward link stronger than the backward, or vice versa. That asymmetry could influence how effectively one term cues the other.

Finally, lexical characteristics such as word frequency may be important. In facts, subjects are often low-frequency words that are less concrete or imageable, especially if unfamiliar, and more difficult to encode than other terms. Predicatives typically contain more familiar, higher-frequency terms (as in definitions; McDaniel et al., 2015). In prior studies of pretesting and fact learning (e.g., Kornell, 2014; Kornell et al., 2009; Metcalfe & Finn, 2011; Storm et al., 2022), the subject has usually been pretested (cf. Kang et al., 2011), leaving the predicative unaddressed. Lexical asymmetries in frequency or familiarity between subject and predicative terms might lead to uneven encoding and retention if one or the other is pretested, potentially reducing any benefits of pretesting for tested or untested terms.

The present study

We investigated how pretesting on different parts of factual statements – specifically, the subject and the predicative – affects memory, relative to studying whole statements. Across nine experiments (including paired experiments

1a/1b and 6a/6b), three primary research questions were addressed:

1. To what extent does pretesting on the subject of a fact or on a key term from the predicative of a fact enhance memory for the tested term?
2. Does the pretesting effect transfer from tested to untested terms within a fact? That is, does pretesting improve memory for untested terms in pretested facts, relative to facts that were only studied?
3. Are pretesting effects for facts influenced by structural factors (e.g., syntax or the presence of additional cue words), semantic features (e.g., associative strength between terms), or lexical characteristics (e.g., word frequency)? Do asymmetries or symmetries in any of these characteristics have an impact?

All experiments shared a common design. First, in an initial learning phase, participants engaged in pretesting or studying of factual statements. With pretesting, either the subject or a key term from the predicative was pretested, whereas with studying, the entire fact was studied. Then, after a 5-min or 24-h retention interval, participants undertook a criterial test, during which memory for the subject or the key term from the predicative of each fact was assessed. As detailed in Fig. 1, pretested facts were assessed on the criterial test for the previously tested term (in the *pretested-same* condition) or for the previously untested term (in the *pretested-different* condition). Studied facts were assessed for the subject or a predicative key term in the *studied* condition.

Experiments 1a–4 employed facts drawn from prior research and varied the terms that were pretested, plus addressed potential effects of syntactic structure. Experiments 5–7 used newly developed facts in which the subject and the relevant key term of the predicative were closely equated in frequency and constrained for associative strength. Those experiments also explored the influence of additional cue words embedded within each fact. By testing on different parts of facts and modifying their properties, we were able to investigate pretesting effects across tested and untested factual materials, while isolating the contributions of structural, semantic, and lexical factors.

Experiments 1a and 1b

Experiments 1a and 1b investigated, respectively, the effects of pretesting on the subject or a key term from the predicative of factual statements, relative to studying intact facts. These initial experiments were designed to explore whether pretesting produces symmetrical learning – benefiting both tested and untested components – or asymmetrical learning, similar to the selective effects typically seen with retrieval practice and word triplets or facts.

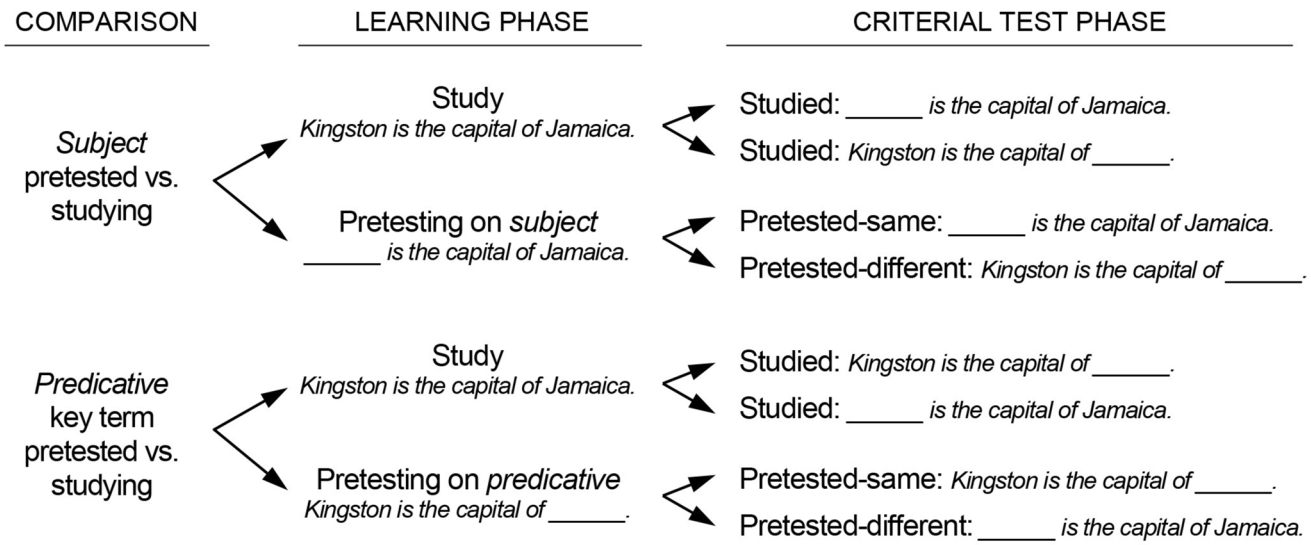


Fig. 1 Learning phase and criterial test conditions. During the learning phase, there were two conditions: in the *study* condition, participants were shown the full factual statement, whereas in the *pretesting* condition, participants were shown an incomplete factual statement and asked to guess either the subject or predicative of the statement. During the criterial test, there were three conditions. Memory for the subject or a key term of the predicative from studied facts was assessed in the *studied* condition, memory for pretested terms from pretested facts was assessed in the *pretested-same* condition, and memory for terms from pretested facts that were not previously tested was assessed in the *pretested-different* condition

Method

Participants

The target sample size for all experiments was determined through an a priori power analysis using G*Power (Faul et al., 2007), which indicated that 34 participants would be sufficient to detect a medium effect size (Cohen's $d = 0.50$) – smaller than the pretesting effects observed by Kornell (2014) for trivia facts – on a dependent-samples t-test at $\alpha = 0.05$ with 80% power. Although the data were ultimately analyzed using a two-way analysis of variance (ANOVA), the power analysis corresponded directly to the main effect of Criterial Test Condition (pretested vs. studied) in that ANOVA. As the study was not formally powered to detect interactions between Criterial Test Condition and Key Term (subject vs. predicative), such interactions should be interpreted cautiously. Each experiment targeted at least 50 participants and we recruited accordingly, although the final sample for Experiment 1b was ultimately slightly smaller due to exclusions.

Fluent English speakers from the UK, USA, Australia, Canada, or New Zealand were recruited via the Prolific Academic platform and compensated US$3 for approximately 20 min of participation. All participants were at least 21 years old, had a Prolific approval rate of 95% or higher, and had completed at least 50 studies. Experiment 1a included 50 participants ($M_{age} = 32.8$ years, 64% female), all of whom completed the study and provided valid data. In Experiment 1b, 50 participants were initially recruited, but four participants were excluded for repeatedly failing to provide guesses for pretest questions, resulting in a final sample of 46 ($M_{age} = 30.6$ years, 78% female). The entire study was conducted with ethics approval and all participants gave informed consent before participating.

Design

Both Experiment 1a and Experiment 1b featured a 3 (Criterial Test Condition: pretested-same vs. pretested-different vs. studied) × 2 (Key Term: subject vs. predicative) within-participants design.

Materials

The stimuli consisted of 32 historical or trivia facts drawn from prior research (Kornell, 2014; Tauber et al., 2013). These facts, which are catalogued in Appendix A, covered various topics such as history, geography, and pop culture, each following a general "*A is B*" structure where *A* is the subject (a single word) and *B* is the predicative phrase. Although the predicative usually consisted of multiple words, one constituent key term was chosen to be tested – specifically, the defining term of the predicative. For example, in the statement "*Kingston is the capital of Jamaica*", "*Kingston*" is the single-word subject, and "*Jamaica*" is the predicative key term from the predicative phrase. In Experiment 1a, half of the statements were presented as incomplete statements with the subject replaced by a blank (e.g., "_____ *is the capital of Jamaica*"), whereas the other half were presented as full statements. In Experiment 1b, half of the statements were also presented as incomplete statements, except that the predicative key term was replaced by a blank (e.g., "*Kingston is the capital of* _____"). The other half of the statements were presented as full statements, as in Experiment 1a.

In prior pretesting research with these materials (e.g., Kornell, 2014), only the subject was typically pretested, usually by presenting the fact as a question (e.g., beginning with "*What*" or "*Which*"). Significant pretesting effects were observed in such instances. The present study differed by examining the effects of pretesting beyond just the subject of each fact, including for tested and untested information, as well as by using a procedure where the facts were presented in fill-in-the-blank rather than question form.

Semantic and structural features Forward and backward associative strengths – semantic features that, in the present case, refer to how strongly the subject evokes the predicative key term and vice versa – were not available for these materials, as such values were not listed in Nelson et al.'s (1998) Free Association Norms. Nonetheless, given the subject-to-predicate nature of the facts and the disparity in word frequencies, associative strength was likely relatively low and asymmetrical. An examination of the facts further revealed that most contained semantic or conceptual asymmetries – that is, one-way relationships where the predicative key term defines the subject but not the reverse. For example, in "*Demeter is the goddess of agriculture,*" "*agriculture*" is a defining characteristic of "*Demeter,*" but "*Demeter*" does not define "*agriculture*" – illustrating a probable directional conceptual link rather than a bidirectional or symmetrical relationship. In some cases, the subject was a proper noun but the key term was not, further contributing to representational asymmetry. Moreover, as previously noted, the predicative itself included additional terms that could function as retrieval cues, marking another structural feature that distinguished predicative key terms from subject terms.

Lexical features Based on frequency measures from Kucera and Francis (1967) – which indicate the number of occurrences per million words in a standard written English corpus – subject word frequencies in the facts ranged from 1 to 129 ($M = 10.1$, $SD = 31.1$; data available for 53% of items),

whereas frequencies for the selected predicative key terms ranged from 1 to 808 ($M = 62.0$, $SD = 162.0$; data available for 78% of items) (Coltheart, 1981). Hence, the key term in the predicative of each fact tended to be a far more common, familiar word. Owing to the high proportion of proper nouns, normed data for concreteness, imageability, and familiarity were available for only 6% of subjects and 41% of predicative key terms in the Scott et al. (2019) Glasgow norms, which include approximately 5,000 English words. Of that normed data, subjects showed numerically lower average ratings (on a 1–7 scale) than key terms for concreteness ($M = 4.9$, $SD = 2.3$ vs. $M = 6.4$, $SD = 0.5$), imageability ($M = 5.5$, $SD = 1.5$ vs. $M = 6.5$, $SD = 0.4$), and familiarity ($M = 5.4$, $SD = 0.3$ vs. $M = 6.0$, $SD = 0.5$), reflecting their greater obscurity and lower vividness.

Procedure

Participants completed a single experimental session on their personal laptop or desktop computer. Each session consisted of three phases: a learning phase, a distractor task, and a criterial test.

Learning phase Participants were first shown instructions explaining that they would encounter two types of statements: incomplete statements with a blank (at the beginning in Experiment 1a; at the end in Experiment 1b), and complete statements to be read. For incomplete statements, they were to guess the missing word. To ensure comprehension, they completed a sample item and two attention checks before proceeding.

Participants then viewed 32 statements presented in random order. Half (16) were presented as pretest trials in incomplete form: the subject (Experiment 1a) or a key term from the predicative (Experiment 1b) was replaced by a blank. Participants typed their guesses within 8 s and were encouraged to guess even if unsure. Immediately afterward, the complete statement appeared for 5 s along with the prompt, "*Read the complete statement, which includes the correct answer*," and the instruction to "*Pay very careful attention to the answer and see how it fits with the rest of the fact*," which was intended to ensure that the feedback was processed properly.

The remaining 16 statements were presented intact for studying, with each statement appearing for 5 s. Participants were instructed to read each statement carefully. The timing of these conditions – 5 s to study statements, versus 8 s to guess plus 5 s of feedback – matched that used in prior pretesting studies (e.g., Kornell et al., 2009; cf. Butowska-Buczyńska et al., 2024), equating exposure time to complete facts (i.e., including the correct answers to any pretest questions).

Distractor task To minimize rehearsal of learned content, participants completed a distractor task involving five 1-min open-ended prompts. For each prompt, they listed as many items as possible from a given category (e.g., celebrities, household items).

Criterial test Participants were tested on all 32 facts from the learning phase. Each fact was presented as an incomplete statement, with a blank replacing either the subject or the key term from the predicative (see Fig. 1). Of the 16 pretested items, eight were tested on the same term that was pretested (e.g., *pretested-same* items), and eight on a different term (e.g., *pretested-different* items). The 16 study-only items were similarly split between subject and predicative testing. The selection of statements that were tested for the subject or for a key term from the predicative was counterbalanced over participants. Participants were required to type a response for every item into a provided textbox, had unlimited time to do so, and once they had finished, were debriefed and compensated.

Scoring

Before formal analysis, participants' criterial test responses were computer scored, with exact matches to the correct answer marked as correct. A rater blind to experimental condition then reviewed all responses and reclassified as correct any answers containing minor misspellings that did not alter meaning (e.g., transposed letters; singular instead of plural), resulting in an additional < 1% of responses being scored as correct across all experiments. Then, in line with established practices in the pretesting literature to control for items that participants likely had pre-existing knowledge of (e.g., Huelser & Metcalfe, 2012; Kornell et al., 2009), any items that were guessed correctly by a participant – that is, during the initial learning phase – were excluded from their data, leading to the removal of 3% of all criterial test data in Experiment 1a and 8% in Experiment 1b. Doing so removed items that participants presumably knew prior to the experiment. This scoring process was repeated in all subsequent experiments, resulting in the removal of approximately 4–6% of all criterial test data.

Analysis plan

We analyzed criterial test performance first by conducting a two-way ANOVA with factors of Criterial Test Condition (pretested-same/-different vs. studied) and Key Term (subject vs. predicative). Criterial Test Condition was collapsed across the two pretesting subtypes to create a balanced two-level factor and to test whether a general pretesting advantage – for retention or transfer – was present across key term types. Where the ANOVA yielded significant effects

Table 1 Pretest performance

Experiment	Mean guessing accuracy	*SE*
1a (*subject* pretested only)	0.07	0.01
1b (*predicative* pretested only)	0.18	0.02
2	0.12	0.02
3	0.09	0.01
4	0.19	0.02
5	0.11	0.07
6a	0.12	0.01
6b	0.12	0.01
7	0.09	0.01

or interactions, we followed up with planned pairwise comparisons conducted separately for subject recall and predicative key term recall. To evaluate the strength of evidence for the null versus the alternative hypothesis, we also calculated Bayes factors (BF_{10}) using the *BayesFactor* package in R (Morey & Rouder, 2012) for all pairwise comparisons. A $BF_{10} > 1$ indicates support for the alternative hypothesis, whereas a $BF_{10} < 1$ favors the null. For interpretability, we report the reciprocal BF_{01} when the evidence favors the null (in practice, BF_{10} is presented for effects supporting the alternative hypothesis, whereas BF_{01} is presented for effects supporting the null).

Results

Pretest performance

The proportion of items guessed correctly on pretest questions administered during the learning phase for Experiments 1a and 1b, as well as for all subsequent experiments, are catalogued in Table 1. As expected, guessing accuracy was generally low (typically ≈ 0.1).

Criterial test performance

Criterial test results for Experiments 1a and 1b are depicted in the top and bottom panels of Fig. 2, respectively.

Experiment 1a A two-way ANOVA on participant-level mean criterial test scores revealed significant main effects of Criterial Test Condition, $F(1, 49) = 18.22$, $MSE = 0.50$, $p < .001$, $\eta_p^2 = 0.27$, and Key Term, $F(1, 49) = 41.66$, $MSE = 1.27$, $p < .001$, $\eta_p^2 = 0.46$, but no significant interaction ($p = .54$). These results reflect better performance on pretested than studied items and for the predicative versus the subject (see Fig. 2, top panel).

Experiment 1b A two-way ANOVA on participant-level mean criterial test scores found a significant main effect of Key Term, $F(1, 45) = 37.29$, $MSE = 1.40$, $p < .001$, $\eta_p^2 = 0.45$, but no significant effect of Criterial Test Condition or a significant interaction (p-values $\geq .11$). These results reflect higher overall performance when the predicative was assessed, but no significant pretesting effects for either key term (see Fig. 2, bottom panel).[1]

Discussion

The results of Experiments 1a and 1b revealed that pretesting on facts with substantial structural, semantic, and lexical asymmetries can yield variable learning for tested and untested terms, with outcomes differing based on the portion of the fact that was pretested. In Experiment 1a, pretesting on the subject enhanced memory for both the subject and a key term from the predicative, illustrating a case where benefits of pretesting for fact learning apply not only to directly pretested material, but also to non-pretested material. In contrast, Experiment 1b showed that pretesting on the predicative evidently did not significantly enhance memory for the tested key term, nor promote transfer to the subject. Together, these findings raise the possibility that the act of making a guess with feedback on a fact may involve different or more strongly engaged cognitive processing of the fact depending on which part is guessed.

Experiment 2

Experiment 2 investigated whether the observed asymmetry in learning outcomes following pretesting in Experiments 1a and 1b would replicate when pretesting occurs on *either* the subject or the predicative across facts, as opposed to only the subject or the predicative. For instance, rather than the predicative being pretested consistently (e.g., "*Kingston is the capital of* _____"; "*Cesium is the softest* _____", etc.), facts in this experiment were pretested on the subject (e.g., "_____ *is the capital of Jamaica*") or the predicative (e.g., "*Cesium is the softest* _____"). This arrangement could reduce the susceptibility to focus on one part of the fact to the exclusion of other parts, as might have occurred in

[1] Prompted by the notably higher guessing accuracy rate in Experiment 1b, we repeated the same two-way ANOVA without excluding correctly guessed items. That analysis revealed a significant main effect of Criterial Test Condition, suggesting that the exclusion of correctly guessed items reduced the observable pretesting effects in this experiment. A comparable ANOVA for Experiment 1a, however, where the guessing success rate was much lower, showed no change in the patterns of results.

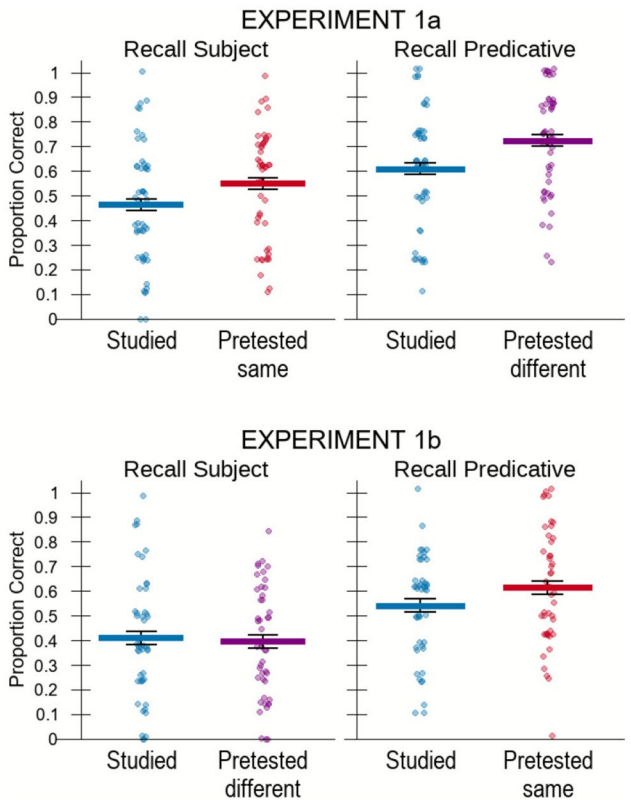


Fig. 2 Experiments 1a and 1b criterial test results. In Experiment 1a, participants were pretested on the subject only, whereas in Experiment 1b, participants were pretested on a predicative key term only. On the criterial test, participants were assessed for recall of the subject and the predicative key term: *pretested-same* indicates that the term being recalled was identical to that which was pretested, whereas *pretested-different* indicates that the term being recalled on the criterial test was different to that which was pretested. The *studied* condition always studied a complete fact beforehand. Error bars indicate standard errors based on the interaction error term of a within-subjects analysis of variance on mean accuracy scores (Loftus & Masson, 1994)

Experiment 1a, and in particular in Experiment1b (in that experiment, participants might have paid attention to the predicative and ignored the subject). Overall, this experiment addressed whether the observed asymmetry in the prior experiments reflected inherent differences in learning with such facts or was a consequence of the guessing procedure.

Method

Participants

We initially recruited 53 participants via Prolific using the same procedures as in prior experiments. Data from three participants were excluded for repeatedly failing to provide guesses for pretest questions, resulting in a final sample of 50 participants (M_{age} = 34.8 years, 58% female) included in analyses.

Design, materials, procedure, scoring, and data analysis

The materials were the same 32 facts used in Experiments 1a and 1b, except that two facts previously guessed correctly by over 30% of participants were replaced with more challenging ones (detailed in Appendix A). During the learning phase, half of the facts were presented as full statements, whereas the other half were incomplete – half missing the subject and half missing a key term from the predicative. On the criterial test, half of the facts that had been pretested were tested on the subject and half on a key term from the predicative; the same split arrangement applied to the facts that had been read in full. The remainder of the criterial test was identical to Experiments 1a and 1b.

As detailed in Fig. 1, items pretested and tested on the same key term (subject or predicative) on the criterial test were classified as *pretested-same* items. Items pretested on one term but tested on the other term were classified as *pretested-different* items. When memory for the subject was assessed, the pretested-same condition refers to the case where the subject had been pretested, whereas the pretested-different condition refers to the case where the predicative had been pretested. When memory for the key term from the predicative was assessed, the pretested-same condition refers to the case where the predicative had been pretested, whereas the pretested-different condition refers to the case where the subject had been pretested.

Results

We performed a 3 (Criterial Test Condition: pretested-same vs. pretested-different vs. studied) × 2 (Key Term: subject vs. predicative) within-participants two-way ANOVA on participant-level mean criterial test scores. That ANOVA yielded a significant main effect of Criterial Test Condition, $F(2, 98) = 3.39$, $MSE = 0.19$, $p = .04$, $\eta_p^2 = 0.07$, and Key Term, $F(1, 49) = 30.85$, $MSE = 1.90$, $p < .001$, $\eta_p^2 = 0.39$, and a significant interaction between Criterial Test Condition and Key Term, $F(2, 98) = 5.43$, $MSE = 0.21$, $p = .01$, $\eta_p^2 = 0.10$. These results, which are depicted in Fig. 3 (top panel), constitute further evidence for asymmetry in fact learning following pretesting, although the pattern is not completely identical to that observed in Experiments 1a or 1b.

We conducted follow-up *t*-tests to explore these patterns. When the subject was pretested, there was no pretesting effect for later recall of the subject (pretested-same vs. studied), $t(49) = 0.35$, $p = .73$, $d = 0.05$, $BF_{01} = 6.14$, but there was strong transfer to the predicative (pretested-different

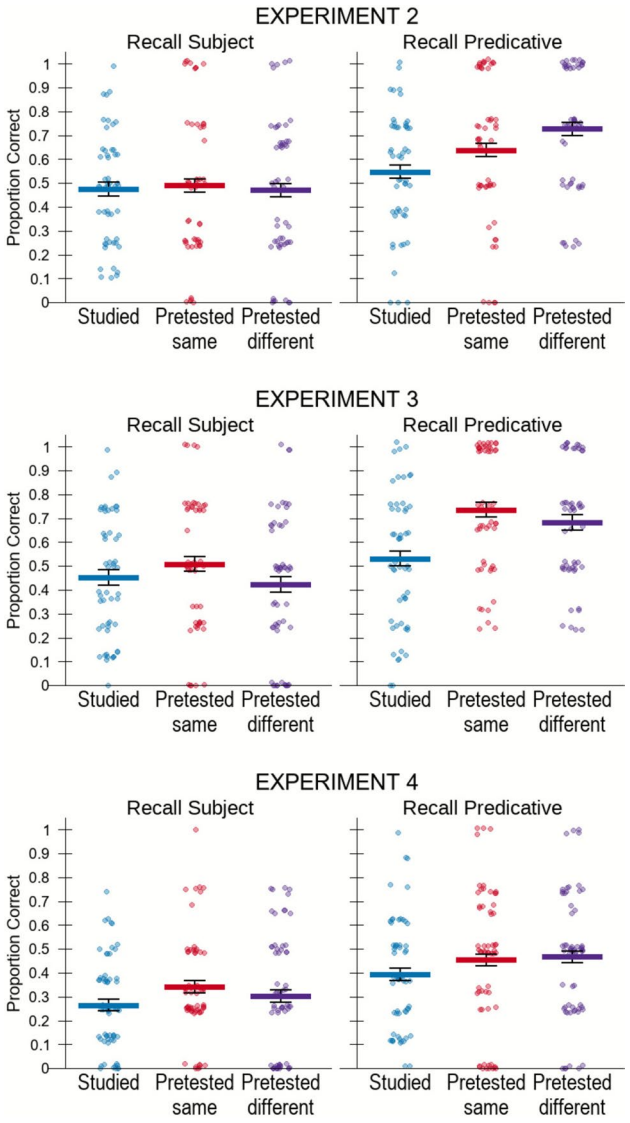


Fig. 3 Experiments 2–4 criterial test results. Relative to Experiment 2, Experiment 3 featured changed sentence syntax and Experiment 4 featured a longer retention interval. On the criterial test, participants were assessed for recall of the subject and the predicative key term: For the case of recalling the subject, *pretested-same* indicates prior pretesting on the subject and *pretested-different* indicates prior pretesting on the predicative. For the case of recalling the predicative, *pretested-same* indicates prior pretesting on the predictive and *pretested-different* indicates prior pretesting on the subject. The studied condition always studied a complete fact beforehand. Error bars indicate standard errors based on the interaction error term of a within-subjects analysis of variance on mean accuracy scores (Loftus & Masson, 1994)

vs. studied), $t(49) = 5.46$, $p < .001$, $d = 0.77$, $BF_{10} = 10{,}931$. When the predicative was pretested, there was no evidence in the Bayes factor analysis for a pretesting effect (pretested-same vs. studied), $t(49) = 1.98$, $p = .05$, $d = 0.28$, $BF_{01} = 1.08$, and no evidence of transfer to the subject, $t(49) = 0.12$, $p = .91$, $d = 0.02$, $BF_{01} = 6.46$.

Discussion

The results of Experiment 2 largely replicate the asymmetrical patterns observed in Experiments 1a and 1b. Specifically, pretesting on the subject produced transfer to the predicative, as seen in Experiment 1a, whereas pretesting on the predicative failed to yield transfer to the subject, as in Experiment 1b. The latter pattern occurred even under a design that presumably reduced the likelihood of participants focusing exclusively on the subject or predicative key term of each fact. The primary divergence was the absence of a pretesting effect for the subject in the pretested-same condition of Experiment 2, contrasting with the benefit observed in Experiment 1a. Remarkably, pretesting in this case enhanced memory for the untested but not the tested term – that is, it improved transfer to related information without boosting recall of practiced content. Together, these findings further suggest that different or more strongly engaged cognitive processing of fact components occurs depending on which part of a fact is guessed.

Experiment 3

To probe the basis of asymmetrical fact learning from pretesting – which could reflect structural, semantic, or lexical factors – Experiment 3 addressed a structural feature, syntactic order, that has been influential in prior memory research (e.g., Wearing, 1971; Weyerts et al., 2002). Whereas the subject always appeared before the predicative in prior experiments – which could have produced a primacy effect for the subject (although it did not always benefit from being pretested) – Experiment 3 featured facts where the predicative always preceded the subject. This modification allowed us to examine whether the previously observed asymmetry arises from syntactic structure or serial position effects.

Method

Participants

We initially recruited 58 participants via Prolific using the same procedures as in prior experiments. Six participants' data were excluded for failing to provide guesses for pretest questions, resulting in a final sample of 52 participants ($M_{age} = 31.7$ years, 60% male).

Design, materials, procedure, scoring, and data analysis

The only change from Experiment 2 was the last rephrasing of each fact such that the predicative always appeared

before the subject in a general "*B is A*" format. For example, "*Kingston is the capital of Jamaica*," was rephrased to "*The capital of Jamaica is Kingston.*"

Results

A two-way ANOVA analogous to that performed for the prior experiment yielded a significant main effect of Criterial Test Condition, $F(2, 102) = 7.88$, $MSE = 0.44$, $p < .001$, $\eta_p^2 = 0.13$, and Key Term, $F(1, 51) = 61.83$, $MSE = 2.78$, $p < .001$, $\eta_p^2 = 0.55$, as well as a significant interaction, $F(2, 102) = 4.69$, $MSE = 0.24$, $p = .01$, $\eta_p^2 = 0.08$. Follow-up tests of the results, shown in Fig. 3 (middle panel), revealed that pretesting on the subject yielded no significant pretesting effect for later recall of the subject (pretested-same vs. studied), $t(51) = 1.31$, $p = .20$, $d = 0.18$, $BF_{01} = 2.96$, along with strong transfer to the predicative (pretested-different vs. studied), $t(51) = 3.40$, $p = .001$, $d = 0.47$, $BF_{10} = 22.20$ – marking another case where pretesting enhanced memory for untested terms only. When the predicative was pretested, there was a pretesting effect for later recall of that term (pretested-same vs. studied), $t(51) = 4.24$, $p < .001$, $d = 0.59$, $BF_{10} = 240$, but no transfer to the subject, $t(51) = 0.69$, $p = .49$, $d = 0.10$, $BF_{01} = 5.27$.

Discussion

The results of Experiment 3 largely replicate those of Experiment 2, including the finding that pretesting on the subject of a fact can enhance memory for untested but not tested terms, although there was stronger evidence of a pretesting effect for the predicative when it was directly pretested (possibly due to it appearing earlier in each fact). Overall, Experiment 3 largely ruled out syntactic structure and serial position effects as explanations for the asymmetry that can occur following pretesting on facts.

Experiment 4

Experiment 4 explored the stability of asymmetrical fact learning over a longer retention interval. In this experiment, participants took the criterial test after at least 24 h had elapsed. Recent work suggests that pretesting effect magnitude can increase with longer retention intervals (Kliegl et al., 2022). If so, then we might expect stronger pretesting effects, perhaps for both the subject and predicative key terms, after 24 h.

Method

This experiment was preregistered at https://aspredicted.org/4mrj-khxr.pdf. All analyses performed for this experiment were preregistered.

Participants

We initially recruited 70 participants via Prolific using the same procedures as the prior experiments, with an added US$1 incentive for completing the second session within 3 h of receiving the link. After excluding seven participants for non-completion, two participants for failing instructions, and one participant for prior knowledge of the materials, the final sample included 60 participants ($M_{age} = 31.6$ years, 53% male).

Design, materials, procedure, scoring, and data analysis

All procedures matched Experiment 2, except that all participants completed two sessions spaced at least 24 h apart and with no intervening distractor task. In the first session, participants passed two attention checks, completed the learning phase, and were reminded to finish the second session on time. The criterial test phase was completed in the second session, accessed via a link sent 24 h later.

Results and discussion

A two-way ANOVA on criterial test results (see Fig. 3, bottom panel) yielded significant main effects of Criterial Test Condition, $F(2, 118) = 3.41$, $MSE = 0.16$, $p = .04$, $\eta_p^2 = 0.06$, and Key Term, $F(1, 59) = 28.86$, $MSE = 1.65$, $p < .001$, $\eta_p^2 = 0.33$. No significant interaction between Criterial Test Condition and Key Term was found ($p = .57$). Follow-up *t*-tests, however, revealed signs of asymmetry: Pretesting on the subject enhanced recall of the subject, $t(59) = 2.46$, $p = .02$, $d = 0.32$, $BF_{10} = 2.24$, with marginal transfer to the predicative, $t(59) = 2.00$, $p = .05$, $d = 0.26$, $BF_{10} = 0.91$, whereas pretesting on the predicative did not significantly improve its later recall, $t(59) = 1.50$, $p = .14$, $BF_{01} = 2.46$, nor enhance recall of the subject, $t(59) = 1.08$, $p = .29$, $BF_{01} = 4.08$. These patterns largely mirror prior experiments, showing that pretesting on the subject of lexically unbalanced facts can enhance its later recall (as in Experiment 1a) and may transfer to the predicative (Experiments 1a–3), whereas pretesting on the predicative of such facts enhances recall of neither the tested term (Experiments 1b–2) nor of the subject (Experiments 1b–3).

Experiment 5

Having ruled out syntactic structure as the main driver of the observed asymmetry – and shown that it persists across retention intervals of at least 24 h – we next turned our focus to the influence of lexical and other structural properties. In the prior experiments, the predicative key terms generally had higher word frequency than the subject terms, plus were accompanied by additional cue words, which may have supported encoding or retrieval. To address these imbalances, Experiment 5 introduced a newly constructed set of facts in which the subject and predicative key terms were lexically balanced – that is, close to evenly matched on word frequency – and were unaccompanied by additional cue words. Associative strength was also constrained but not fully equated. If lexical factors are determinative, then these balancing efforts should reduce or eliminate asymmetrical fact learning.

Method

This experiment was preregistered at https://aspredicted.org/hv9m-hx5r.pdf. All analyses performed for this experiment were preregistered, except where noted as exploratory.

Participants

We initially recruited 50 participants via Prolific using the same procedures as in prior experiments. Data from three participants were excluded for repeatedly failing to provide guesses for pretest questions, yielding a final sample of 47 participants (M_{age} = 33.5 years, 56% male) included in analyses.

Design, materials, procedure, scoring, and data analysis

This experiment closely resembled that of Experiment 2 in nearly all respects, including procedures, scoring, and data analysis. The sole exception was the materials, which consisted of 32 newly constructed trivia facts (see Appendix B). These facts originated as word pairs from Nelson et al. (1998) that were chosen to have comparable word frequency between terms and to meet associative strength constraints. From these pairs, facts were created by inserting a verb between the cue and target to form complete sentences without any additional words in a general "*A is B*" format (e.g., the pair *Advil–aches* became "*Advil is for aches*"). Hence, unlike the materials used previously, these facts had subject and predicative key terms that were much more closely matched in word frequency and other lexical characteristics, met associative strength constraints, and contained no additional cues besides a verb that linked the two key terms.

Semantic and structural features Forward associative strength was kept weak (ranging from 0.010 to 0.014, M = 0.013, SD = 0.0014), whereas backward associative strength was generally very low (ranging from 0 to 0.015, M = 0.0014, SD = 0.0043). Due to limitations in available word pairs, perfect symmetry in associative strength between subject and predicative key terms was not achievable (it was especially difficult to identify sufficient pairs with weak backward associations). Hence, as in prior experiments, these facts were asymmetrical in terms of forward and backward associative strength.

Inspection of the facts further revealed that conceptual or semantic asymmetry also remained present. For example, in the fact "*Incense produces aroma*," *aroma* is a property of *incense*, but the reverse is not true. Such asymmetries reflect real-world directional relationships – such as cause and effect or object and attribute. Unlike the materials used in earlier experiments, however – which featured numerous proper nouns – only two out of 32 subjects in the present set were proper names, thereby reducing potential influences related to familiarity with specific individuals or entities. The lack of additional cue words for the predicative also served to eliminate that source of structural asymmetry.

Lexical features The subject and predicative key terms were very closely matched on word frequency (data available for 100% of terms), ranging from 0 to 216 for cues (M = 15.7, SD = 41.5) and targets (M = 16.5, SD = 41.9), with a maximum frequency difference between paired words of 7 (M = 0.8, SD = 4.1) (Nelson et al., 1998). This close matching helped control for lexical differences between terms. Normed data for concreteness, imageability, and familiarity from Scott et al. (2019) were available for 60% of subjects and 63% of predicative key terms; the mean ratings were more comparable between the subject and predicative key terms for concreteness (M = 4.9, SD = 1.3 vs. M = 5.2, SD = 1.5), imageability (M = 5.3, SD = 1.3 vs. M = 5.5, SD = 1.3), and familiarity (M = 5.2, SD = 0.8 vs. M = 5.6, SD = 0.7) than in the prior experiments, corresponding to the greater similarity in word frequency, although ratings remained numerically slightly higher for predicative key terms. In addition, for the 28% of facts with available ratings for *both* terms, the average differences between subject and predicative terms were small: familiarity (M_{diff} < 0.1, SD = 0.5), imageability (M_{diff} = 0.1, SD = 1.1), and concreteness (M_{diff} = 0.1, SD = 1.1).

Results

A two-way ANOVA analogous to that performed for Experiments 2 and 3 yielded a significant main effect of Criterial Test Condition, $F(2, 92) = 61.78$, $MSE = 2.49$, $p < .001$,

$\eta_p^2 = 0.57$, no significant main effect of Key Term ($p = .09$), and a significant interaction between Criterial Test Condition and Key Term, $F(2, 92) = 3.86$, $MSE = 0.11$, $p = .02$, $\eta_p^2 = 0.08$. The lack of a significant main effect of Key Term differs from the prior experiments. Follow-up *t*-tests further confirmed that when the subject was pretested, there was a pretesting effect for later recall of the subject (pretested-same vs. studied), $t(46) = 11.22$, $p < .001$, $d = 1.64$, $BF_{10} > 1$ million, as well as strong transfer to the predicative (pretested-different vs. studied), $t(46) = 5.20$ $p < .001$, $d = 0.76$, $BF_{10} = 4{,}029$. Similarly, when the predicative was pretested, there was a pretesting effect for later recall of the predicative (pretested-same vs. studied), $t(46) = 6.02$, $p < .001$, $d = 0.88$, $BF_{10} = 56{,}740$, as well as strong transfer to the subject (pretested-different vs. studied), $t(46) = 7.74$, $p < .001$, $d = 1.13$, $BF_{10} > 1$ million.

For insights into the significant interaction, we conducted exploratory follow-up *t*-tests comparing Key Term performance (subject vs. predicative) within each level of Criterial Test Condition. There were no significant differences in recall of the subject versus the predicative for the pretested-same ($t(46) = 0.05$, $p = .96$, $BF_{01} = 6.31$) or pretested-different conditions ($t(46) = 0.21$, $p = .84$, $BF_{01} = 6.19$), whereas performance was significantly lower for recall of the subject than the predicative in the studied condition, $t(46) = 3.37$, $p = .01$, $BF_{10} = 19.71$. These patterns are depicted in Fig. 4 (top panel), in which it is evident that disparate performance across key terms in the studied condition contributed to the significant interaction.

Discussion

The results of Experiment 5 reveal that when word frequency and other lexical features are closely equated between the subject and predicative key terms – and when additional cue words are removed – pretesting improves learning of both tested and untested parts of facts, producing comparable benefits despite associative and semantic asymmetries still being present. Mechanistically, these findings suggest that the earlier asymmetries likely stemmed from lexical factors. When lexical imbalances are minimized, guessing appears to strengthen associative links across fact elements more uniformly, supporting memory for both tested and untested components.

Experiments 6a and 6b

Experiments 6a and 6b attempted to replicate the symmetrical learning patterns observed with lexically balanced facts in Experiment 5, plus examined whether additional cue words – which were present in earlier experiments but absent in Experiment 5 – might affect the results. Both experiments used the same facts as Experiment 5 but reintroduced a single cue word per fact, referencing either the subject or the predicative key term. Doing so offered potential insights into whether additional cue words might have contributed to the asymmetries observed in Experiments 1a–3. Overall, the length and syntactic structure of the facts in Experiments 6a–6b more closely resembled the facts used in Experiments 1a–2 than the facts used in Experiment 5, which were, by design, highly reminiscent of paired associates. If cue words selectively enhance learning for specific terms, then we would expect asymmetrical pretesting effects to re-emerge; conversely, if lexical balance is the primary determinant, then pretesting should continue to produce roughly equal benefits across tested and untested terms.

Method

Both experiments were preregistered at https://aspredicted.org/qs3t-5jhc.pdf. All analyses performed for both experiments were preregistered.

Participants

We initially recruited 54 and 53 participants, respectively, for Experiments 6a and 6b via Prolific, using the same procedures as in prior experiments. In each experiment, two participants were excluded for repeatedly failing to provide guesses for pretest questions, resulting in final samples of 52 for Experiment 6a ($M_{age} = 32.0$ years, 52% female) and 51 for Experiment 6b ($M_{age} = 33.2$ years, 47% female).

Design, materials, procedure, scoring, and data analysis

Both experiments closely resembled Experiment 5, including drawing from the same factual materials, except that two items with a high guessing rate were replaced or modified with more challenging terms (see Appendix B). Importantly, each fact now featured an additional cue for the subject (Experiment 6a) or predicative key term (Experiment 6b). For instance, the fact "*Advil is for aches*" was modified to "*Advil is the medicine for aches*" (added cue word referring to subject; Experiment 6a) and "*Advil cures joint aches*" (added cue word referring to predicative key term; Experiment 6b). These cue words, again drawn from the Nelson et al. (1998) database, were all plausible modifiers or descriptors for the associated subject or predicative key term. They were also matched in average frequency across the two experiments (data available for 100% of terms) and tended to be higher frequency words than the corresponding subject or predicative key term (Experiment 6a: ranging from 2 to 200, $M = 60.9$, $SD = 60.1$; Experiment 6b: ranging from 2 to 239, $M = 70.9$, $SD = 62.9$) (Nelson et al., 1998). Association strength data were limited, but available

measures indicate that these words had low forward associative strength with their respective targets (subject cues: $M = 0.032$; data available for 6% of terms; predicative cues: $M = 0.018$, data available for one term), consistent with their role as plausible but minimally pre-associated cues.

Results

Criterial test results for Experiments 6a and 6b are depicted in the middle panels of Fig. 4.

Experiment 6a

A two-way ANOVA on criterial test results yielded a significant main effect of Criterial Test Condition, $F(2, 102) = 49.95$, $MSE = 1.93$, $p < .001$, $\eta_p^2 = 0.57$, a significant main effect of Key Term, $F(1, 50) = 5.76$, $MSE = 0.14$, $p = .02$, $\eta_p^2 = 0.10$, and no significant interaction ($p = .18$). Follow-up t-tests revealed that when the subject was pretested, there was a strong pretesting effect for later recall of the subject, $t(51) = 6.49$, $p < .001$, $d = 0.90$, $BF_{10} = 386{,}061$, and strong transfer to the predicative, $t(51) = 4.39$, $p < .001$, $d = 0.61$, $BF_{10} = 383$. When the predicative was pretested, there was a strong pretesting effect for later recall of the predicative key term, $t(51) = 7.75$, $p < .001$, $d = 1.07$, $BF_{10} > 1$ million, and strong transfer to the subject, $t(51) = 6.62$, $p < .001$, $d = 0.92$, $BF_{10} = 598{,}385$. These results strongly resemble patterns observed in Experiment 5, with an exception being better overall recall of subject terms.

Experiment 6b

A two-way ANOVA on criterial test results yielded a significant main effect of Criterial Test Condition, $F(2, 100) = 66.80$, $MSE = 2.47$, $p < .001$, $\eta_p^2 = 0.57$, no significant main effect of Key Term, $F(1, 46) = 3.09$, $MSE = 0.09$, $p = .09$, $\eta_p^2 = 0.06$, and no significant interaction ($p = .32$). Follow-up t-tests revealed that when the subject was pretested, there was a strong pretesting effect for later recall of the subject, $t(50) = 7.70$, $p < .001$, $d = 1.08$, $BF_{10} > 1$ million, and strong transfer to the predicative, $t(50) = 7.90$, $p < .001$, $d = 1.11$, $BF_{10} > 1$ million. When the predicative was pretested, there was a strong pretesting effect for later recall of the predicative key term, $t(50) = 8.80$, $p < .001$, $d = 1.23$, $BF_{10} > 1$ million, and strong transfer to the subject, $t(50) = 6.68$, $p < .001$, $d = 0.94$, $BF_{10} = 682{,}623$. These results also closely resemble patterns observed in Experiment 5.

Discussion

The results of Experiments 6a and 6b conceptually replicated the primary result of Experiment 5, showing consistently symmetrical pretesting effects on both directly tested and untested terms when lexical features were carefully controlled. Moreover, these experiments rule out the presence of extra cue words as a sufficient condition for producing asymmetrical learning through pretesting. Hence, the presence of additional cue words – and presumably any greater attention or cuing that such words might have conferred on the respective subject or predicative term – likely was not the main driver of the asymmetrical learning patterns observed in Experiments 1a–3. Rather, lexical factors appear to critically shape the differential engagement of cognitive operations that can produce asymmetrical learning.

Experiment 7

Mirroring the approach taken with Experiment 4, the final experiment explored the stability of symmetrical fact learning over a longer retention interval of at least 24 h. If symmetrical pretesting benefits observed for lexically balanced facts are robust, then we would expect comparable improvements for both tested and untested terms even after a delay.

Method

This experiment was preregistered at https://aspredicted.org/c26p-7gy7.pdf. All analyses performed for this experiment were preregistered.

Participants

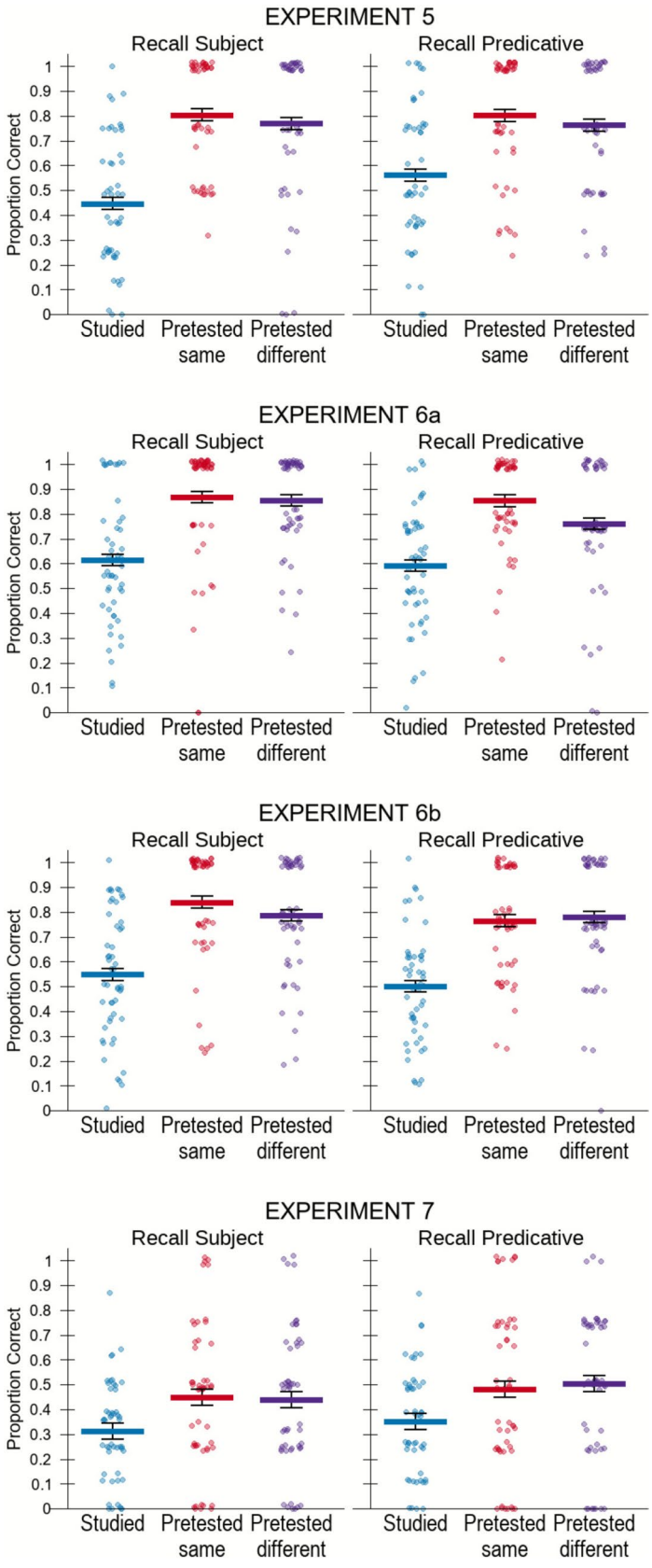
We initially recruited 51 participants using the same procedures as Experiment 4. Data from one participant were excluded as the participant did not complete the full experiment. As a result, data from 50 participants ($M_{age} = 32.9$ years, 62% male) were analyzed.

Design, materials, procedure, scoring, and data analysis

All aspects were based on that of Experiment 5 with the primary exception that the criterial test occurred in a separate session. The delayed test design was based on Experiment 4.

Results and discussion

A two-way ANOVA on criterial test results (see Fig. 4, bottom panel) yielded a significant main effect of Criterial Test Condition, $F(2, 98) = 10.63$, $MSE = 0.62$, $p < .001$, $\eta_p^2 = 0.18$, no significant main effect of Key Term, $F(1, 49) = 3.02$, $MSE = 0.15$, $p = .09$, $\eta_p^2 = 0.06$. and no significant interaction ($p = .80$). Follow-up t-tests revealed that when the subject was pretested, there was a pretesting effect for later recall of the subject, $t(49) = 3.20$, $p < .01$, $d = 0.45$, $BF_{10} = 12.90$, and


EXPERIMENT 5
Recall Subject
Recall Predicative
EXPERIMENT 6a
Recall Subject
Recall Predicative
EXPERIMENT 6b
Recall Subject
Recall Predicative
EXPERIMENT 7
Recall Subject
Recall Predicative
Proportion Correct
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
Studied
Pretested same
Pretested different

◂**Fig. 4** Experiments 5–7 criterial test results. Relative to Experiment 4, Experiments 6a/6b featured additional cue words and Experiment 7 featured a longer retention interval. On the criterial test, participants were assessed for recall of the subject and the predicative key term: For the case of recalling the subject, *pretested-same* indicates prior pretesting on the subject and *pretested-different* indicates prior pretesting on the predicative. For the case of recalling the predicative, *pretested-same* indicates prior pretesting on the predictive and *pretested-different* indicates prior pretesting on the subject. The studied condition always studied a complete fact beforehand. Error bars indicate standard errors based on the interaction error term of a within-subjects analysis of variance on mean accuracy scores (Loftus & Masson, 1994)

significant transfer to the predicative, $t(49) = 3.75$, $p < .001$, $d = 0.53$, $BF_{10} = 56.72$. When the predicative was pretested, there was a pretesting effect for later recall of the predicative key term, $t(49) = 2.93$, $p < .01$, $d = 0.41$, $BF_{10} = 6.68$, and evidence of transfer to the subject, $t(49) = 2.95$, $p < .01$, $d = 0.42$, $BF_{10} = 7.07$. Overall, as in Experiments 5 and 6b, these results demonstrate symmetrical fact learning following pretesting with lexically balanced facts.

General discussion

Although prior studies have demonstrated benefits of pretesting on the subject term of historical or trivia facts (e.g., Kornell et al., 2009), which might imply that pretesting generally enhances fact learning, the present experiments reveal that the magnitude and extent of pretesting effects for facts, including for tested and untested content, vary considerably depending on which term is pretested and how learning is assessed. Experiments 1a–4 drew on the same facts used in prior pretesting research and found that pretesting can produce highly asymmetrical learning. Pretesting on the subject term tended to benefit the predicative term more than the subject term itself, whereas pretesting on the predicative improved memory mainly for the tested term. In Experiments 5–7, however, which featured a new set of facts in which the subject and predicative terms were closely matched in word frequency and other lexical characteristics, pretesting yielded highly symmetrical learning, enhancing memory for both terms regardless of which was tested – a pattern that resembles that observed in Pan et al. (2019) and, across experiments, in Hays et al. (2013).

The entire set of results, which are summarized in Table 2, provides answers for all three primary research questions that guided this study. First, the extent to which pretesting on the subject or the predicative of a fact enhances memory for that fact is variable. Reasons for those patterns are considered further below. Second, the pretesting effect can transfer successfully from tested to untested terms within a fact – and sometimes even manifests only for untested and not tested terms. Those findings extend pretesting and fact learning research to encompass transfer effects, plus contrast with findings showing limited transfer and high specificity following retrieval practice on facts (e.g., Hinze & Wiley, 2011; Pan et al., 2016a, 2016b). Third, pretesting effects for facts are strongly influenced by lexical characteristics such as word frequency and possibly concreteness, imageability, and familiarity. Semantic features such as associative strength and conceptual relatedness, as well as structural factors such as syntax, appear to play limited roles, at least when it comes to memory for tested and untested subject and predicative key terms. Crucially, although associative and semantic asymmetries between subject and predicative terms were present across all experiments, manipulations intended to reduce or modify those asymmetries – such as altering syntax (Experiment 3) or adding cues (Experiments 6a and 6b) – had little discernible impact. Although prior work shows that unrelated word pairs often fail to benefit from pretesting, our findings imply that once minimal semantic relatedness is present, variations in associative or semantic strength do not affect pretesting outcomes to the same extent as lexical factors.

Factors that might have been considered pivotal to pretesting effects for fact learning were also addressed. For instance, the semantic richness of factual statements (Kornell, 2014) can no longer be regarded as a sufficient condition for generating pretesting effects for either tested or untested content. The provision of immediate feedback is likewise insufficient. Stimulus complexity also does not by itself determine pretesting effectiveness or constrain transfer, as evidenced by comparable results across materials that were more complex (Experiments 6a and 6b) or simpler (Experiment 5) in terms of length, syntactic structure, and informational content. Ultimately, lexical characteristics emerged as the most robust determinant of pretesting outcomes – a conclusion supported by the clear divide in results for lexically unbalanced (Experiments 1a–3) versus balanced (Experiments 4–6b) materials.

A process account of pretesting effects for tested and untested fact terms

We next turn to the cognitive processes underlying the four pretesting scenarios examined in this study. These scenarios involve pretesting on either the subject or predicative term of a fact, under conditions where these terms differ in

relative word frequency – being either lexically unbalanced, with one term higher in frequency than the other, or lexically balanced, where both terms were low in frequency but closely equated and usually not proper nouns. This account assumes that, during initial guessing, pretesting engages semantic more than episodic processes, with learners drawing on semantic memory to interpret the cue and generate plausible target responses.

Pretesting with lexically unbalanced facts

In the first scenario, pretesting on the subject of a lexically unbalanced fact, a participant is pretested with a question such as, "_____ *is the capital of Jamaica*," which includes the predicative cue "*Jamaica*," a term with which they are likely somewhat familiar (i.e., they recognize this country). While attempting to guess, they actively encode and process this cue, drawing on their knowledge of it to generate plausible answers (e.g., various cities in the Caribbean) but the correct answer is unknown (despite some knowledge of the predicative cue) and the guess is incorrect. When feedback is provided, participants have the opportunity to encode the correct answer ("*Kingston*") and relate it to the predicative term.

Under this scenario, memory for the predicative term – which is actively recognized and processed during guessing and subsequently linked with the subject during feedback – is enhanced, improving subsequent recall of that term compared to a study-only condition. In contrast, the subject term, which was not successfully guessed and is learned only during feedback, is more obscure and thus more difficult to encode. Although its presentation during feedback can enable it to be linked with the already-familiar predicative term, both the encoding of the subject and the strength of this association are constrained by the subject's low lexical accessibility. Consequently, memory for the subject term shows only a modest benefit at best, relative to a study-only condition.

In the second scenario, pretesting on the predicative of a lexically unbalanced fact, a participant is pretested with a question such as, "*Kingston is the capital of* _____," which includes the subject cue "*Kingston*" – a highly unfamiliar, low-frequency proper noun. Although this term is likely processed to some extent, its limited semantic associations and low imageability make it a weak cue. Consequently, the participant is unlikely to use it effectively to generate plausible guesses, instead relying on general knowledge or schema (e.g., thinking about countries with capital cities). The correct predicative term ("*Jamaica*") is not retrieved successfully and is instead only learned through feedback. Encoding of that term at that point also remains limited, no stronger than what would be expected under a study-only condition.

Under this scenario, the subject term is unlikely to have been processed thoroughly during the act of guessing due to its low lexical accessibility and, although presumably linked with the predicative term during feedback, shows little-to-no memory advantage. Although the predicative term is more recognizable, it is learned only through feedback and forms a weak association with the obscure subject, yielding a modest memory benefit at best. Consequently, only a pretesting effect for the predicative term is likely to occur, if at all.

Table 2 Summary of criterial test results

Experiment and stimulus features	After subject pretested, is there a statistically significant pretesting effect for…		After predicative term pretested, is there a statistically significant pretesting effect for…	
	Tested term (*subject*)?	Untested term (*predicative*)?	Tested term (*predicative*)?	Untested term (*subject*)?
Experiments 1a–4 (asymmetrical, structural, semantic, lexical features)				
1a (*subject* pretested only)	Yes ($d = 0.49$)	Yes ($d = 0.41$)	-	-
1b (*predicative* pretested only)	-	-	No ($d = 0.28$)	No ($d < 0.10$)
2	No ($d < 0.10$)	Yes ($d = 0.77$)	No ($d = 0.28$)	No ($d < 0.10$)
3	No ($d < 0.20$)	Yes ($d = 0.47$)	Yes ($d = 0.59$)	No ($d < 0.10$)
4	Yes ($d = 0.32$)	No ($d = 0.26$)	No ($d < 0.20$)	No ($d < 0.20$)
Experiments 5–7 (symmetrical lexical; asymmetrical structural, semantic features)				
5	Yes ($d = 1.64$)	Yes ($d = 0.76$)	Yes ($d = 0.88$)	Yes ($d = 1.13$)
6a	Yes ($d = 0.90$)	Yes ($d = 0.61$)	Yes ($d = 1.07$)	Yes ($d = 0.92$)
6b	Yes ($d = 1.08$)	Yes ($d = 1.11$)	Yes ($d = 1.23$)	Yes ($d = 0.94$)
7	Yes ($d = 0.45$)	Yes ($d = 0.53$)	Yes ($d = 0.41$)	Yes ($d = 0.42$)

Experiment 3 introduced variations in syntax and Experiments 6a–6b restored additional cues

Both scenarios are consistent with the asymmetrical patterns observed in Experiments 1b–3. The absence of pretesting effects for lexically inaccessible terms also echoes findings with paired associates involving obscure or foreign words, where pretesting yields no memory enhancement on criterial tests involving cued recall (e.g., Potts & Shanks, 2014; Seabrooke et al., 2021). Exceptions, however, emerged in Experiments 1a and 4, where pretesting on the subject yielded a pretesting effect for the subject. These patterns suggest that lexically obscure terms are not inherently immune to pretesting benefits but may require attentional biases or strategic encoding behaviors, or may simply become more apparent after longer retention intervals due to differential forgetting relative to studied conditions (Kliegl et al., 2022). The results of Experiment 1a also mirror prior studies which have shown pretesting effects for low-frequency subject terms when only the subject is tested (e.g., Kornell, 2014; Kornell et al., 2009).

Pretesting with lexically balanced facts

In the third scenario, pretesting on the subject of a lexically balanced fact, a participant is presented with a question such as "_____ *climbs trees*," which includes the predicative cue "*trees*," a recognizable term. While guessing, they actively process this cue, drawing on their knowledge of it to generate plausible answers (e.g., various animals), thus strengthening memory for the predicative term. Even if they guess incorrectly, the correct answer might be mentally activated, and upon feedback presentation, participants encode it ("*Bears*") and link it to the predicative term. Crucially, lexical obscurity is not a factor for either the subject or the predicative term. Memory for the predicative term – which was engaged during guessing and likely connected to the subject during feedback – is reliably enhanced, improving recall versus a study-only condition. Memory for the subject term is also strengthened, having been activated during guessing, easily encoded with feedback, and/or readily associated with the familiar predicative term. Hence, both terms show strong memory benefits, producing symmetrical learning patterns.

In the fourth scenario, pretesting on the predicative term of a lexically balanced fact, a participant is presented with a question such as "*Bears climb* _____." Given the comparable (albeit low) lexical accessibility of the subject and predicative terms, this fourth scenario operates almost identically to the third; only the positions of the subject and predicative terms are swapped. The subject ("*Bears*") is actively processed during guessing, and although the correct predicative term ("*trees*") may not be retrieved, it is subsequently encoded during feedback and linked to the lexically accessible subject. Consequently, both terms benefit from pretesting, and a symmetrical pattern of memory enhancement again occurs.

Broader theoretical implications

Although the present results do not necessarily favor any specific theoretical account of the pretesting effect, they underscore the importance of lexical constraints in shaping the various mechanisms that have been proposed. For example, if mediator generation drives the effect (Kornell et al., 2009), then the lexical accessibility of the cues is crucial. Even low-frequency terms, as in Experiments 5–7, can support mediator generation if they are sufficiently accessible (i.e., recognizable), whereas more obscure terms, such as the proper nouns in Experiments 1a–4, are more likely to impede that process. Similarly, in the search set activation account (Grimaldi & Karpicke, 2012), lexical accessibility may shape the range of candidate answers that are activated during guessing, consistent with the finding that word frequency influences the speed and ease of searching one's mental lexicon (Forster & Chambers, 1973). Lexical factors might also interact with memory-for-change or recursive reminding processes (Wahlheim & Jacoby, 2013; see also Hintzman, 2011), affecting how strongly cues and targets are integrated into memory, with more accessible elements yielding stronger integration.

With respect to attentional theories, the present findings challenge two claims. First, in the context of pretesting with immediate correct answer feedback, it is not necessarily the case that attention is directed only towards the target (Potts & Shanks, 2014); there are evidently cases where pretesting also increases attention to cues as well (Experiments 5–7; see also Hays et al., 2013; Pan et al., 2019), potentially even during the act of processing feedback. Second, it is also evident that attention alone does not govern transfer; even when it is presumably directed to both cue and target, transfer may only occur when both terms are similarly accessible, suggesting that attentional processes must interact with cue familiarity or other factors. Relatedly, although methodological and stimulus differences complicate interpretation, the asymmetrical learning patterns that commonly occur with pretesting on text passages might similarly reflect inherent lexical differences across text components.

The present findings also raise the possibility that the robust pretesting effects for both tested and untested

information observed in Hays et al. (2013) and Pan et al. (2019) were facilitated or supported by the lexical properties of their materials (e.g., *frog–pond*; *gift*, *rose*, *wine*), which appear to have a strong degree of lexical accessibility. If so, then the similarity in results between those studies and Experiments 5–7 suggests that simple, non-ecological materials can nonetheless reveal a core characteristic of the pretesting effect – that is, similar effects for tested and untested materials under lexically balanced conditions – with a meaningful degree of generalizability (which validates an investigative approach that begins with simpler materials). Moreover, the process account outlined here could be interpreted in light of classic transfer theories that stress overlap between conditions of encoding and retrieval (e.g., Thorndike & Woodworth, 1901). Specifically, by enabling easier processing and stronger integration of cue–target pairs, lexically accessible materials may promote such overlap, thus facilitating learning of both tested and untested information.

Limitations and future directions

Although the present experiments went beyond prior research by testing memory for both subject and predicative key terms, the effects of pretesting for other components such as additional cue words or verbs remains to be addressed. Future research could explore whether similar patterns occur with term-definition facts (e.g., Pan & Rickard, 2017) or more complex fact learning (e.g., Kang et al., 2011). A fill-in-the-blank format was consistently used for both pretesting and criterial tests, leaving room to examine alternative test formats (e.g., short answer, multiple choice) and variations in feedback implementation. Other procedural or design elements (e.g., instructions, exposure durations, number of practice trials or items, presentation formats, and participant cognitive abilities (Pan, Yu, et al., 2025) also merit further consideration.

Lexical accessibility indexed by word frequency showed strong effects, but further investigation could clarify how lexical properties interact with other factors such as encoding strategies. Although semantic and structural influences appeared limited – and associative strength was constrained in Experiments 5–7 (but not tested under symmetrical conditions) – these factors along with other semantic and structural variables warrant further study. Additionally, normed lexical and semantic data were available for only a subset of terms in Experiments 1a–4, which limits stimulus characterization. Using more fully normed materials may improve precision but could reduce ecological validity.

Potential links to prior work on associative symmetry and asymmetry (Kahana, 2002) could also be explored in future studies. Much of that prior work has used simple cue-target pairs and cued recall practice following initial study (e.g., Sarkis & Montag, 2021), which differs substantially from the materials and pretesting procedures used here. These methodological differences complicate direct comparison, highlighting the challenge facing research bridging these approaches.

Conclusions

The present experiments reveal that the magnitude of pretesting effects for facts is substantially impacted by the terms that are pretested and their lexical accessibility. Lexical characteristics – particularly word frequency – are strong influences, outweighing semantic, associative, or structural factors. Lexically accessible cues are more easily learned through pretesting and can support transferable learning, whereas low-accessibility terms constrain such learning. From a pedagogical standpoint, strategically considering which terms are pretested and which parts of a fact are most important to learn are likely to help maximize the capacity of pretesting to enhance fact learning.

Appendix A

Table 3 Fact stimuli from Experiments 1a–4

No	Facts from Experiments 1a, 1b, 2, and 4	Facts from Experiment 3
1	*Ahab is the captain of Pequod*	*The captain of Pequod is Ahab*
2	*Baikal is the deepest lake*	*The deepest lake is Baikal*
3	*Bamboo is the tallest grass*	*The tallest grass is Bamboo*
4	*Canberra is the capital of Australia*	*The capital of Zimbabwe is Harare*
5	*Cesium is the softest metal*	*The softest metal is Cesium*
6	*Corsica is the birthplace of Napoleon*	*The birthplace of Napoleon is Corsica*
7	*Demeter is the goddess of agriculture*	*The goddess of agriculture is Demeter*
8	*Douglas is the actor of Spartacus*	*The actor of Spartacus is Douglas*
9	*Garnet is the birthstone for January*	*The birthstone for January is Garnet*
10	*Guernica is painted by Picasso*	*The painting by Picasso is Guernica*
11	*Hakarl is made from sharks*	*The food made from sharks is Hakarl*
12	*Kingston is the capital of Jamaica*	*The capital of Jamaica is Kingston*
13	*Larry is the mascot of Twitter*	*The mascot of Twitter is Larry*
14	*Max is the Grinch's dog*	*The Grinch's dog is Max*
15	*Redwood is the tallest tree*	*The tallest tree is Redwood*
16	*Segar is the creator of Popeye*	*The creator of Popeye is Segar*
17	*Alfred is the butler of Batman*	*The girlfriend of Batman is Selina*
18	*Balsa is the lightest wood*	*The lightest wood is Balsa*
19	*Bologna is the oldest university*	*The oldest university is Bologna*
20	*Carcross is the smallest desert*	*The smallest desert is Carcross*
21	*Cookie is the oldest parrot*	*The oldest parrot is Cookie*
22	*Delaware is the first state*	*The first state is Delaware*
23	*Dodoma is the capital of Tanzania*	*The capital of Tanzania is Dodoma*
24	*Frankfort is the capital of Kentucky*	*The capital of Kentucky is Frankfort*
25	*George is the father of Barbie*	*The father of Barbie is George*
26	*Haggis is made from sheep*	*The pudding made from sheep is Haggis*
27	*Hannibal is a general of Carthage*	*The Roman general from Carthage is Hannibal*
28	*Kip is the currency of Laos*	*The currency of Laos is Kip*
29	*Malbork is the largest castle*	*The largest castle is Malbork*
30	*Melody is the daughter of Ariel*	*The daughter of Ariel is Melody*
31	*Scoville is the unit of spiciness*	*The unit of spiciness is Scoville*
32	*Stapes is the smallest bone*	*The smallest bone is Stapes*
33*	*Harare is the capital of Zimbabwe*	-
34*	*Selina is the girlfriend of Batman*	-

Subject and predicative key terms are underlined. Items 33 and 34 replaced items 4 and 17 from Experiment 2 onwards

Appendix B

Table 4 Fact stimuli from Experiments 5–7

No	Facts from Experiments 5 and 7	Facts from Experiment 6a	Facts from Experiment 6b
1	*Advil cures aches*	*Advil is the medicine for aches*	*Advil cures joint aches*
2	*Allergies cause hives*	*Allergies are conditions that cause hives*	*Allergies cause itchy hives*
3	*Bears climb trees*	*Blots are a type of stain*	*Blots are annoying stains*
4	*Calculators show digits*	*Calculators are devices that show digits*	*Calculators show numerous digits*
5	*Camels have lumps*	*Camels are animals with lumps*	*Camels have fat lumps*
6	*Chaos causes frustration*	*Chaos is a situation that causes frustration*	*Chaos causes serious frustration*
7	*Choirs sing hymns*	*Choirs have singers who sing hymns*	*Choirs sing beautiful hymns*
8	*Cigars produce ash*	*Cigars are products that produce ash*	*Cigars produce grey ash*
9	*Clowns have makeup*	*Clowns are artists with makeup*	*Clowns have heavy makeup*
10	*Cobras are serpents*	*Cobras are reptiles that are serpents*	*Cobras are venomous serpents*
11	*Cubes are prisms*	*Cubes are shapes that are prisms*	*Cubes are square prisms*
12	*Dance to be happy*	*Dance is a sport to be happy*	*Dance to be extra happy*
13	*Depletion produces scarcity*	*Depletion is a circumstance that produces scarcity*	*Depletion produces severe scarcity*
14	*Dragons are monsters*	*Dragons are beasts that are monsters*	*Dragons are fierce monsters*
15	*Essays are prose*	*Essays are text that are prose*	*Essays are short prose*
16	*Fever is an illness*	*Fever is a symptom of an illness*	*Fever is a minor illness*
17	*Fraud is deception*	*Fraud is a move of deception*	*Fraud is criminal deception*
18	*Geese can honk*	*Geese are birds that honk*	*Geese can produce honks*
19	*Incense produces aroma*	*Incense are materials that produce aroma*	*Incense produces warm aroma*
20	*Litter causes pollution*	*Litter is waste that causes pollution*	*Litter causes terrible pollution*
21	*Oregano is a herb*	*Oregano is a herb that complements garlic*	*Oregano complements the vegetable garlic*
22	*Otters are mammals*	*Otters are a type of mammal*	*Otters are wild mammals*
23	*Parties have music*	*Parties are events with music*	*Parties have loud music*
24	*Pearls are jewels*	*Pearls are objects that are jewels*	*Pearls are precious jewels*
25	*Pepsi contains caffeine*	*Pepsi is a drink that contains caffeine*	*Pepsi contains low caffeine*
26	*Possums have pouches*	*Possums are creatures with pouches*	*Possums have baby pouches*
27	*Saxophones are woodwinds*	*Saxophones are instruments that are woodwinds*	*Saxophones are popular woodwinds*
28	*Taboos are restrictions*	*Taboos are acknowledged as restrictions*	*Taboos are moral restrictions*
29	*Telescopes can magnify*	*Telescopes are tools that magnify*	*Telescopes can powerfully magnify*
30	*Tigers are predators*	*Tigers are cats that are predators*	*Tigers are apex predators*
31	*Toothpaste removes plaque*	*Toothpaste is a substance that removes plaque*	*Toothpaste removes stubborn plaque*
32	*Vagrants are jobless*	*Vagrants are individuals who are jobless*	*Vagrants are usually jobless*

Subject and predicative key terms are underlined. Items 3 and 21 were replaced or modified for Experiments 6a and 6b

Acknowledgements Thanks to Bridget McConnell, Kean Hsu, and Tim Rickard for helpful comments on an earlier draft of this manuscript.

Authors' contributions S. C. Pan led the conceptualization, methodology, writing, visualization, supervision, funding acquisition, and project administration. He also provided supporting contributions to formal analysis and resources.

F. A. Siauw provided supporting contributions to conceptualization and methodology. She led software development, validation, formal analysis, investigation, resources, and data curation. She also provided supporting contributions to the original draft writing.

Funding This research was supported by a Faculty of Arts and Social Sciences (FASS) grant awarded to S. C. Pan and a Thesis Support Fund grant awarded to F. A. Siauw.

Data availability Materials, data, and analysis code are available at: https://osf.io/epq5w/

Code availability Available at: https://osf.io/epq5w/

Declarations

Conflicts of interest/Competing interests The authors declare no conflicts of interest.

Ethics approval Ethics approval for the entire study was obtained before data collection (NUS-Psych-DERC #2022-September-14).

Consent to participate All participants were treated in accordance with the Declaration of Helsinki, and informed consent was obtained prior to participation.

Consent for publication Not applicable.

Open Practices Statement The materials, data, and analysis code for this study are publicly available via the Open Science Framework (OSF) at: https://osf.io/epq5w/

Experiments 4 (https://aspredicted.org/4mrj-khxr.pdf), 5 (https://aspredicted.org/hv9m-hx5r.pdf) 6a and 6b (https://aspredicted.org/qs3t-5jhc.pdf), and 7 (https://aspredicted.org/c26p-7gy7.pdf) were preregistered.

References

Bartl, J., Kliegl, O., & Bäuml, K.-H.T. (2024). The role of mediators for the pretesting effect. *Memory, 0*(0), 1–11. https://doi.org/10.1080/09658211.2024.2323930

Butler, A. C. (2010). Repeated testing produces superior transfer of learning relative to repeated studying. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 36*(5), 1118–1133. https://doi.org/10.1037/a0019902

Butowska-Buczyńska, E., Hanczakowski, M., & Zawadzka, K. (2024). Errorful learning of trivia questions and answers: The role of study time. *Memory & Cognition*. https://doi.org/10.3758/s13421-024-01608-6

Carpenter, S. K. (2011). Semantic information activated during retrieval contributes to later retention: Support for the mediator effectiveness hypothesis of the testing effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 37*(6), 1547–1552. https://doi.org/10.1037/a0024140

Carpenter, S. K., Pashler, H., & Vul, E. (2006). What types of learning are enhanced by a cued recall test? *Psychonomic Bulletin & Review, 13*(5), 826–830. https://doi.org/10.3758/BF03194004

Collins, A., & Loftus, E. (1975). A spreading activation theory of semantic processing. *Psychological Review, 82*, 407–428. https://doi.org/10.1037/0033-295X.82.6.407

Coltheart, M. (1981). The MRC psycholinguistic database. *The Quarterly Journal of Experimental Psychology, 33A*(4), 497–505. https://doi.org/10.1080/14640748108400805

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*, 175–191. https://doi.org/10.3758/BF03193146

Forster, K. I., & Chambers, S. M. (1973). Lexical access and naming time. *Journal of Verbal Learning and Verbal Behavior, 12*(6), 627–635. https://doi.org/10.1016/S0022-5371(73)80042-8

Geller, J., Carpenter, S. K., Lamm, M. H., Rahman, S., Armstrong, P. I., & Coffman, C. R. (2017). Prequestions do not enhance the benefits of retrieval in a STEM classroom. *Cognitive Research: Principles and Implications, 2*(1), 42. https://doi.org/10.1186/s41235-017-0078-z

Grimaldi, P. J., & Karpicke, J. D. (2012). When and why do retrieval attempts enhance subsequent encoding? *Memory & Cognition, 40*(4), 505–513. https://doi.org/10.3758/s13421-011-0174-0

Hays, M. J., Kornell, N., & Bjork, R. A. (2013). When and why a failed test potentiates the effectiveness of subsequent study. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 39*(1), 290–296. https://doi.org/10.1037/a0028468

Hintzman, D. L. (2011). Research strategy in the study of memory: Fads, fallacies, and the search for the "coordinates of truth." *Perspectives on Psychological Science, 6*(3), 253–271. https://doi.org/10.1177/1745691611406924

Hinze, S. R., & Wiley, J. (2011). Testing the limits of testing effects using completion tests. *Memory, 19*(3), 290–304. https://doi.org/10.1080/09658211.2011.560121

Huelser, B. J., & Metcalfe, J. (2012). Making related errors facilitates learning, but learners do not know it. *Memory & Cognition, 40*(4), 514–527. https://doi.org/10.3758/s13421-011-0167-z

Jacoby, L. L., & Wahlheim, C. N. (2013). On the importance of looking back: The role of recursive remindings in recency judgments and cued recall. *Memory & Cognition, 41*(5), 625–637. https://doi.org/10.3758/s13421-013-0298-5

Jacoby, L. L., Wahlheim, C. N., & Coane, J. H. (2010). Test-enhanced learning of natural concepts: Effects on recognition memory, classification, and metacognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 36*(6), 1441. https://doi.org/10.1037/a0020636

James, K. K., & Storm, B. C. (2019). Beyond the pretesting effect: What happens to the information that is not pretested? *Journal of Experimental Psychology: Applied, 25*(4), 576–587. https://doi.org/10.1037/xap0000231

Kahana, M. J. (2002). Associative symmetry and memory theory. *Memory & Cognition, 30*(6), 823–840. https://doi.org/10.3758/BF03195769

Kang, S. H. K., Pashler, H., Cepeda, N. J., Rohrer, D., Carpenter, S. K., & Mozer, M. C. (2011). Does incorrect guessing impair fact learning? *Journal of Educational Psychology, 103*(1), 48–59. https://doi.org/10.1037/a0021977

Kasahara, K., & Yanagisawa, A. (2024). Learning new verbs with known cue words: The relative effects of noun and adverb cues. *Language Teaching Research, 28*(1), 138–155. https://doi.org/10.1177/13621688211001613

Kliegl, O., Bartl, J., & Bäuml, K.-H. T. (2022). The pretesting effect comes to full fruition after prolonged retention interval. *Journal of Applied Research in Memory and Cognition*.https://doi.org/10.1037/mac0000085

Knight, J. B., Hunter Ball, B., Brewer, G. A., DeWitt, M. R., & Marsh, R. L. (2012). Testing unsuccessfully: A specification of the underlying mechanisms supporting its influence on retention. *Journal of*

Memory and Language, 66(4), 731–746. https://doi.org/10.1016/j.jml.2011.12.008

Kornell, N. (2014). Attempting to answer a meaningful question enhances subsequent learning even when feedback is delayed. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 40*(1), 106–114. https://doi.org/10.1037/a0033699

Kornell, N., Hays, M. J., & Bjork, R. A. (2009). Unsuccessful retrieval attempts enhance subsequent learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 35*(4), 989–998. https://doi.org/10.1037/a0015729

Kornell, N., & Vaughn, K. E. (2016). How Retrieval Attempts Affect Learning. In *Psychology of Learning and Motivation* (Vol. 65, pp. 183–215). Elsevier. https://doi.org/10.1016/bs.plm.2016.03.003

Kucera, H., & Francis, W. N. (1967). *Computational analysis of present-day American English*. Brown University Press.

Little, J. L., & Bjork, E. L. (2016). Multiple-choice pretesting potentiates learning of related information. *Memory & Cognition, 44*(7), 1085–1101. https://doi.org/10.3758/s13421-016-0621-z

Loftus, G. R., & Masson, M. E. (1994). Using confidence intervals in within-subject designs. *Psychonomic Bulletin & Review, 1*(4), 476–490. https://doi.org/10.3758/BF03210951

Lockhart, R. S. (1969). Retrieval asymmetry in the recall of adjectives and nouns. *Journal of Experimental Psychology, 79*(1, Pt.1), 12–17. https://doi.org/10.1037/h0026944

McDaniel, M. A., Bugg, J. M., Liu, Y., & Brick, J. (2015). When does the test-study-test sequence optimize learning and retention? *Journal of Experimental Psychology: Applied, 21*(4), 370–382. https://doi.org/10.1037/xap0000063

Mera, Y., Gabriel Rodriguez, & Marin-Garcia, E. (2021). Unraveling the benefits of experiencing errors during learning: Definition, modulating factors, and explanatory theories. *Psychonomic Bulletin & Review*, 13.

Metcalfe, J. (2017). Learning from errors. *Annual Review of Psychology, 68*(1), 465–489. https://doi.org/10.1146/annurev-psych-010416-044022

Metcalfe, J., & Finn, B. (2011). People's hypercorrection of high-confidence errors: Did they know it all along? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 37*(2), 437–448. https://doi.org/10.1037/a0021962

Metcalfe, J., & Huelser, B. J. (2020). Learning from errors is attributable to episodic recollection rather than semantic mediation. *Neuropsychologia, 138*, 107296. https://doi.org/10.1016/j.neuropsychologia.2019.107296

Morey, R. D., & Rouder, J. N. (2012). *BayesFactor: Computation of Bayes Factors for Common Designs* (p. 0.9.12–4.7) . https://doi.org/10.32614/CRAN.package.BayesFactor

Nelson, D. L., McEvoy, C. L., & Schreiber, T. A. (1998). The University of South Florida word association, rhyme, and word fragment norms. http://www.usf.edu/FreeAssociation/.

Pan, S. C., & Carpenter, S. K. (2023). Prequestioning and pretesting effects: A review of empirical research, theoretical perspectives, and implications for educational practice. *Educational Psychology Review, 35*(4), 97. https://doi.org/10.1007/s10648-023-09814-5

Pan, S. C., Gopal, A., & Rickard, T. C. (2016a). Testing with feedback yields potent, but piecewise, learning of history and biology facts. *Journal of Educational Psychology, 108*(4), 563–575. https://doi.org/10.1037/edu0000074

Pan, S. C., Lovelett, J., Stoeckenius, D., & Rickard, T. C. (2019). Conditions of highly specific learning through cued recall. *Psychonomic Bulletin & Review, 26*(2), 634–640. https://doi.org/10.3758/s13423-019-01593-x

Pan, S. C., & Rickard, T. C. (2017). Does retrieval practice enhance learning and transfer relative to restudy for term-definition facts? *Journal of Experimental Psychology: Applied, 23*(3), 278–292. https://doi.org/10.1037/xap0000124

Pan, S. C., & Rickard, T. C. (2018). Transfer of test-enhanced learning: Meta-analytic review and synthesis. *Psychological Bulletin, 144*(7), 710–756. https://doi.org/10.1037/bul0000151

Pan, S. C., & Sana, F. (2021). Pretesting versus posttesting: Comparing the pedagogical benefits of errorful generation and retrieval practice. *Journal of Experimental Psychology: Applied*. https://doi.org/10.1037/xap0000345

Pan, S. C., Wong, C. M., Potter, Z. E., Mejia, J., & Rickard, T. C. (2016b). Does test-enhanced learning transfer for triple associates? *Memory & Cognition, 44*(1), 24–36. https://doi.org/10.3758/s13421-015-0547-x

Pan, S. C., Yu, L., Wong, M. J., Selvarajan, G., & Teo, A. Z. (2025). Do individual differences in working memory capacity, episodic memory ability, or fluid intelligence moderate the pretesting effect? *Journal of Memory and Language, 142*, Article 104608. https://doi.org/10.1016/j.jml.2025.104608

Potts, R., & Shanks, D. R. (2014). The benefit of generating errors during learning. *Journal of Experimental Psychology: General, 143*(2), 644–667. https://doi.org/10.1037/a0033194

Richland, L. E., Kornell, N., & Kao, L. S. (2009). The pretesting effect: Do unsuccessful retrieval attempts enhance learning? *Journal of Experimental Psychology: Applied, 15*(3), 243–257. https://doi.org/10.1037/a0016496

Rickard, T. C., Healy, A. F., & Bourne, L. E. (1994). On the cognitive structure of basic arithmetic skills: Operation, order, and symbol transfer effects. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20*(5), 1139–1153. https://doi.org/10.1037/0278-7393.20.5.1139

Sana, F., & Carpenter, S. K. (2023). Broader benefits of the pretesting effect: Placement matters. *Psychonomic Bulletin & Review*. https://doi.org/10.3758/s13423-023-02274-6

Sarkis, J. T., & Montag, J. L. (2021). The effect of lexical accessibility on Spanish-English intra-sentential codeswitching. *Memory & Cognition, 49*(1), 163–180. https://doi.org/10.3758/s13421-020-01069-7

Schlichting, M. L., Mumford, J. A., & Preston, A. R. (2015). Learning-related representational changes reveal dissociable integration and separation signatures in the hippocampus and prefrontal cortex. *Nature Communications, 6*(1), 8151. https://doi.org/10.1038/ncomms9151

Scott, G. G., Keitel, A., Becirspahic, M., Yao, B., & Sereno, S. C. (2019). The Glasgow norms: Ratings of 5,500 words on nine scales. *Behavior Research Methods, 51*(3), 1258–1270. https://doi.org/10.3758/s13428-018-1099-3

Seabrooke, T., Mitchell, C. J., Wills, A. J., & Hollins, T. J. (2021). Pretesting boosts recognition, but not cued recall, of targets from unrelated word pairs. *Psychonomic Bulletin & Review, 28*(1), 268–273. https://doi.org/10.3758/s13423-020-01810-y

St. Hilaire, K. J., Chan, J. C. K., & Ahn, D. (2024). Guessing as a learning intervention: A meta-analytic review of the prequestion effect. *Psychonomic Bulletin & Review, 31*(2), 411–441. https://doi.org/10.3758/s13423-023-02353-8

Storm, B. C., James, K. K., & Stone, S. M. (2022). Pretesting can be beneficial even when using the internet to answer questions. *Memory, 30*(4), 388–395. https://doi.org/10.1080/09658211.2020.1863990

Tauber, S. K., Dunlosky, J., Rawson, K. A., Rhodes, M. G., & Sitzman, D. M. (2013). General knowledge norms: Updated and expanded from the Nelson and Narens (1980) norms. *Behavior Research Methods, 45*(4), 1115–1143. https://doi.org/10.3758/s13428-012-0307-9

Thorndike, E. L., & Woodworth, R. S. (1901). The influence of improvement in one mental function upon the efficiency of other functions. II. The estimation of magnitudes. *Psychological Review, 8*(4), 384–395. https://doi.org/10.1037/h0071280

Tulving, E., & Thomson, D. M. (1973). Encoding specificity and retrieval processes in episodic memory. *Psychological Review, 80*(5), 352–373. https://doi.org/10.1037/h0020071

Vaughn, K. E., & Rawson, K. A. (2012). When is guessing incorrectly better than studying for enhancing memory? *Psychonomic Bulletin & Review, 19*(5), 899–905. https://doi.org/10.3758/s13423-012-0276-0

Wahlheim, C. N., & Jacoby, L. L. (2013). Remembering change: The critical role of recursive remindings in proactive effects of memory. *Memory & Cognition, 41*(1), 1–15. https://doi.org/10.3758/s13421-012-0246-9

Wahlheim, C. N., Smith, W. G., & Delaney, P. F. (2019). Reminders can enhance or impair episodic memory updating: A memory-for-change perspective. *Memory, 27*(6), 849–867. https://doi.org/10.1080/09658211.2019.1582677

Wearing, A. J. (1971). Word class and serial position in the immediate recall of sentences. *Psychonomic Science, 25*(6), 338–340. https://doi.org/10.3758/BF03335898

Weyerts, H., Penke, M., Münte, T. F., Heinze, H.-J., & Clahsen, H. (2002). Word order in sentence processing: An experimental study of verb placement in German. *Journal of Psycholinguistic Research*. https://doi.org/10.1023/a:1015588012457

Zawadzka, K., Zaborowska, O., Butowska, E., Piątkowski, K., & Hanczakowski, M. (2023). Guessing can benefit memory for related word pairs even when feedback is delayed. *Memory & Cognition*. https://doi.org/10.3758/s13421-022-01385-0