**Article:**

Corps, R.E. orcid.org/0000-0001-6027-8109 and Meyer, A.S. (Accepted: 2025) Multiple repetitions lead to the long-term elimination of the word frequency effect. Journal of Experimental Psychology: Learning Memory and Cognition. ISSN 0278-7393 (In Press)

https://doi.org/10.1037/xlm0001486

**Multiple repetitions lead to the long-term elimination of the word frequency effect**

Ruth E. Corps[1, 2] & Antje S. Meyer[2, 3]

[1] School of Psychology, University of Sheffield

[2] Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands

[3] Donders Centre for Brain, Cognition and Behavior , Radboud University, Nijmegen, The Netherlands

Please address correspondence to:

Ruth Elizabeth Corps

School of Psychology

219 Portobello, Broomhall

University of Sheffield

Sheffield, S1 4DP

United Kingdom

r.corps@sheffield.ac.uk

Authors' accepted manuscript – in press in *Journal of Experimental Psychology : Learning, Memory and Cognition.*

**Abstract**

Current theories of speaking suggest that the structure of the lexicon is flexible and changes with exposure. We tested this claim in two experiments that investigated whether the word frequency effect was moderated by item repetition within and across experimental sessions. Participants named high frequency (HF) and low frequency (LF) pictures (Experiment 1) and words (Experiment 2) six times. In both experiments, participants were faster to name HF than LF pictures or words, but this effect was eliminated with repetition. Importantly, this word frequency effect was still absent when participants returned up to two weeks later and named old HF and LF pictures, whose names they had produced before, together with new HF and LF pictures, whose names they had not produced. These findings suggest that producing a word multiple times in short succession alters its long-term accessibility, making it easier to produce later.

Keywords: Word frequency, repetition priming, lexical access

## Introduction

One of the most robust effects in the word processing and production literature is the word frequency effect: More frequent words (e.g., *dog*) tend to be recognised and produced faster than less frequent words (e.g., *stag*; Monsell, Doyle, & Haggard, 1989; Oldfield & Wingfield, 1965). Word frequency is a corpus-based variable that is typically defined as the number of occurrences of a word per one million words in corpora of texts, such as books, newspapers, or television/movie subtitles (e.g., Brysbaert, Mandera, & Keuleers, 2018). How can this corpus-based variable affect the efficiency of lexical access in individual speakers and listeners? A straightforward explanation of word frequency effects is that corpus frequencies roughly correspond to the frequencies with which individual speakers encounter words, and so words that are encountered frequently are more accessible and easier to recognise and produce than words that are less frequently encountered. Thus, the word frequency effect results from cumulative learning from repeated encounters of words. This view is consistent with current theories that suggest that the structure of the mental lexicon is flexible (e.g., Damian & Als, 2005; Howard, Nickels, Coltheart, & Cole-Virtue, 2006; Oppenheim, Dell, & Schwartz, 2010).

This account of the word frequency effect implies that a trace is left in memory when a person encounters a word, and that the trace is durable so that a more and more accessible representation of a word is built up over time (e.g., Estes & Maddox, 2002). Consistent with this assumption, there is a large literature on repetition priming showing that speakers benefit from repetition and, for example, name words or pictures faster when they are shown for a second time (for a review see Francis, 2014). However, most studies of repetition priming have concerned short-term changes in a word's accessibility. If memory traces are to lead to long-term word frequency effects, they must be durable enough to support cumulative learning when words are repeated after longer time intervals. Whether these memory traces are durable enough has rarely been studied. Our study addressed this gap using a naming task to elicit words that were likely to be known by participants, but that varied in word frequency. To test whether individual encounters with a word affected a participant's ease of accessing it later, we investigated how the word frequency effect was moderated by item repetition within an experimental session and, more importantly, across sessions.

Our study builds on earlier work examining the interaction between repetition and word frequency effects. In a seminal study, Jescheniak and Levelt (1994) familiarised participants with a set of pictures with high frequency (HF) and low frequency (LF) names and then asked them to name the pictures three times. As in previous research, participants were faster to name HF (mean (*M*) = 649 ms) than LF pictures (*M* = 711 ms). Importantly, this word frequency effect was stable across three repetitions of the picture. Similarly, Levelt, Praamstra, Meyer, Helenius, and Salmelin (1998) found a stable word frequency effect of around 40 ms across 12 presentations during picture naming in a pilot study.

But other studies have found that pictures with LF names benefitted from repetition more than those with HF names (e.g., Francis, Gurrola, & Martínez, 2022; Griffin & Bock, 1998; La Heij, Puerta-Melguizo, van Oostrum, & Starreveld, 1999). For example, Corps and Meyer (2023) conducted a conceptual replication of Jescheniak and Levelt's study, using different materials, and a direct replication using the same materials. In both experiments, participants were faster to name pictures with HF than LF names, replicating Jescheniak and Levelt. Importantly, however, there was an interaction between word frequency and repetition: The word frequency effect was larger on the picture's first presentation than on the second or third presentation. Similarly, Gollan, Montoya, Fennema-Notestine, and Morris (2005, Experiment 2; see also Gollan, Montoya, Cera, & Sandoval, 2008) found that bilinguals, who use words in each language less frequently, benefitted from repetition during picture naming more than monolinguals, who will have used the picture names more frequently. Thus, the frequency-of-use difference between bilinguals and monolinguals was reduced by repetition. Importantly, and consistent with Corps and Meyer, repetition effects were largest between the first and second presentation of the pictures. The interaction between repetition and word frequency is important because word frequency is a lexical (rather than conceptual or visual) variable, and so the interaction suggests that recent naming affects the ease of accessing the picture's name for later naming. In other words, the interaction suggests that encountering words alters their future accessibility, particularly when words are less frequent. But although previous studies have investigated the interaction between repetition and word frequency effects, they have not assessed whether naming has long-term consequences for a word's accessibility, as required for cumulative learning.

We tested this issue in a study similar to Corps and Meyer (2023). Participants named pictures (Experiment 1) or words (Experiment 2) six times in a single experimental session. We expected to replicate the interaction between repetition and the word frequency effect, which would suggest that even when words are well known, recent use can affect their accessibility in a specific context. In addition, we assessed the persistence of repetition-related reductions in the word frequency effect by testing participants again up to two weeks later. If the word frequency effect is still absent or reduced, then it would suggest that production of well-known but infrequent words affects their long-term accessibility. Such a finding would support the view that the word frequency effect arises as a result of accumulating experience with words.

Participants in Experiment 1 repeatedly named pictures, and so needed to identify the pictured object, retrieve its name, and articulate this name (e.g., Johnson, Paivio, & Clark, 1996). Research suggests that much of the variability in picture naming arises during picture recognition (e.g., Vitkovitch & Tyrell, 1995) and that recognition strongly benefits from repetition (e.g., Francis & Sáenz, 2007; Francis, Corral, Jones, & Sáenz, 2008; Lachman,

Lachman, Thronesbery, & Sala, 1980). Although we closely matched the pictures in the HF and LF conditions for conceptual and visual characteristics, it is likely that at least part of the main effect of repetition in Experiment 1 arose during visual and conceptual processing, rather than at the lexical level.

To address this issue, participants in Experiment 2 produced words without their accompanying pictures. After naming each word (e.g., "dog"), participants judged whether a second word (e.g., "hairy") indicated a property of the concept denoted by the word that they had just named. We included both of these tasks to encourage participants to activate both the form and the meaning of the words, given that research suggests that the word frequency effect is located at multiple levels of lexical access (e.g., Kittredge, Dell, Verkuilen, & Schwartz, 2008). Participants named the pictures for the first time in a second session, and so we investigated whether the production of the picture names (without the accompanying pictures) in the first session reduced the word frequency effect for picture names (with the pictures) in the second session.

Experiment 2 is similar in design to earlier studies investigating the longevity of repetition priming effects in bilingual speakers. Francis and Sáenz (2007) examined repetition priming in picture naming after a short delay (within a session) and a long delay (between two sessions, one week apart). At encoding, participants either named pictures in one language (Spanish or English) or translated a word from Spanish to English or vice versa. At test, participants then named the pictures in either Spanish or English. Regardless of the task used at encoding, repetition priming facilitated picture naming after a one-week delay. Thus, using the picture name, without the presentation of the picture itself, facilitated later picture naming.

However, Francis and Sáenz (2007) did not vary the word frequency of the picture names, and so it is possible that the repetition effect at least partly arose at the conceptual level. In another study, Tsuboi, Francis, and Jameson (2021) varied the word frequency of the picture names, but had participants read the picture names or classify them as either natural or manufactured during encoding. They found a repetition priming effect, which was stronger for LF than HF items in the same experimental session, but no repetition priming effect was seen after a one-week delay. However, participants named or classified each word only once, and so it is possible that they did not have enough exposure to the picture names to lead to long-term repetition priming effects.

In sum, we do not know whether production of well-known but infrequent words affects their longer-term accessibility. Our study addressed this issue using a naming task in two online experiments. In Experiment 1, participants named HF and LF pictures six times in a single session. They then returned and named the same HF and LF pictures together with new HF and LF pictures in a second (Session 2; 3-5 nights after Session 1) and third session (Session 3; 10-14 nights after Session 2). Thus, we assessed the persistence of any repetition-related reduction in the word frequency effect. Given that repetition could benefit visual and conceptual, rather than lexical, processing, we also conducted Experiment 2, in which participants produced the HF and LF words without their accompanying pictures in Session 1 and then returned and named the pictures for the first time in a second session (Session 3; 7-10 nights after Session 1).

# Experiment 1

## Method

### *Transparency and openness*

We report how we determined our sample size, all data exclusions (if any), all manipulations, and all measures in the following sections. All data, analysis code, and list of stimuli are available at https://osf.io/pa8tb/. This study's design and its analysis were not pre-registered.

*Participants*

Our sample size was based on Corps and Meyer (2023), who found an interaction between word frequency and repetition with 40 participants. We recruited 45 native Dutch speakers from the Max Planck Institute for Psycholinguistics participant database (33 females, 10 males, 2 NA; *M*age = 23 years) who participated in exchange for 20 euros. All participants lived in The Netherlands and reported no known speaking, reading, or hearing impairments. Ethical approval for the study was given by the Ethics Board of the Social Sciences Faculty at Radboud University. We discarded data from three participants for Session 1 due to technical errors, and so the analysis for this session is based on data from 42 participants (32 females, 8 males, 2 NA; *M*age = 24 years). One participant provided only five trials for Session 2 and a further three participants dropped out, and so the analysis for Session 2 is based on the data from 38 participants (29 females, 7 males, 2 NA; *M*age = 24 years) and Session 3 is based on the data from 39 participants (28 females, 7 males, 2 NA; *M*age = 23 years).

*Materials*

For Session 1, we selected 64 pictures from the Dutch Bank of Standardised Stimuli (BOSS; Decuyper, Brysbaert, Brodeur, & Meyer, 2021; https://osf.io/kwu87/), which is a database of coloured photographs of everyday objects. All descriptive statistics for the objects were taken from the BOSS database. Half of the pictures had HF names, while the other half had LF names, and the two conditions differed in SUBTLEX frequency estimates ($t(62) = 7.17$, $p < .001$; see Table 1). They also differed in Zipf frequency ($t(62) = 18.35$, $p < .001$), which is a logarithmic scale ranging from 1 (very low frequency words; frequencies of 1 per 100 million words) to 6 (very high frequency content words) or 7 (function words, pronouns, and verb forms like "have"; Van Heuven, Mandera, Keuleers, & Brysbaert, 2014). Zipf frequency is derived from the SUBTLEX frequency estimates.

We also matched the two sets of words for object agreement ($t(62) = -0.64$, $p = .52$), which is a five-point rating of how well participants thought the picture represented its actual concept, and name agreement ($t(62) = -0.32$, $p = .75$), which is the percentage of participants who agree on a particular (dominant) name for the picture. The conditions were also matched for age of acquisition ($t(58) = -1.64$, $p = .11$). Where possible, we matched the word onsets of the picture names in the HF and LF conditions to ensure any differences in response latencies could not be attributed to differences in detecting word onset. Twenty-five of the word onsets in the HF condition were matched to word onsets in the LF condition.

Participants named the pictures for the first time in Session 1 and then again in Sessions 2 and 3. To ensure that any frequency effects for these items did not disappear simply because participants were bored of these items, we selected an additional 64 pictures from the BOSS database to serve as new items. Just like the old items, half of the new items had HF names while the other half had LF names. The two conditions differed in SUBTLEX ($t(62) = 10.82$, $p < .001$) and Zipf frequency estimates ($t(62) = 19.83$, $p < .001$). Importantly, SUBTLEX ($F(1, 124) = 2.99$, $p = .09$) and Zipf frequency estimates ($F(1, 124) = 2.56$, $p = .11$) did not differ for the old and new items, and there was no interaction between frequency and item type for either SUBTLEX ($F(1, 124) = 2.78$, $p = .10$) or Zipf frequency estimates ($F(1, 124) = 0.11$, $p = .74$).

We again matched the word onsets of the picture names in the HF and LF conditions. Nineteen of the word onsets in the HF condition were matched to the word onsets in the LF condition. We also matched the new HF and LF items for object agreement ($t(62) = -0.63$, $p =$

.50) and name agreement ($t(62) = 0.91$, $p = .37$). But as more suitable items were not available in the database, the new HF and LF items differed age of acquisition ($t(47) = -4.88$, $p < .001$). This means that a word frequency effect for the new items may in part be due to a difference between HF and LF items in age of acquisition. Given that the focus of our study was primarily the stability of the frequency effect for old items (and not whether a frequency effect was present for new items), this potential confound does not undermine our argument.

Table 1. Maximum, minimum, means, and standard deviations (SD) of frequency measures (SUBTLEX and Zipf), object agreement, name agreement (%), and age of acquisition for high and low frequency items presented in Session 1, 2, and 3 (old items) or Session 2 and 3 only (new items).

| | High-frequency old | | | | Low-frequency old | | | | High-frequency new | | | | Low-frequency new | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Max | Min | Mean | SD | Max | Min | Mean | SD | Max | Min | Mean | SD | Max | Min | Mean | SD |
| SUBTLEX | 156.92 | 11.71 | 41.56 | 31.82 | 2.97 | 0.11 | 1.23 | 0.88 | 83.63 | 13.19 | 30.91 | 15.59 | 2.93 | 0.14 | 1.04 | 0.94 |
| Zipf | 5.20 | 4.07 | 4.52 | 0.29 | 3.48 | 2.14 | 2.96 | 0.38 | 4.92 | 4.12 | 4.45 | 0.19 | 3.47 | 2.20 | 2.85 | 0.41 |
| Object agreement | 4.82 | 3.39 | 4.40 | 0.35 | 4.92 | 3.72 | 4.45 | 0.31 | 4.83 | 3.55 | 4.35 | 0.39 | 4.86 | 3.52 | 4.40 | 0.33 |
| Name agreement | 100 | 84 | 93.53 | 4.98 | 100 | 82 | 93.97 | 5.89 | 100 | 80 | 93.31 | 6.66 | 100 | 81 | 91.75 | 7.04 |
| Age of acquisition | 8.59 | 3.95 | 5.90 | 1.25 | 7.50 | 4.99 | 6.34 | 0.74 | 7.01 | 4.01 | 5.33 | 0.85 | 7.39 | 5.33 | 6.46 | 0.61 |

*Design*

Each stimulus was presented to participants six times in Session 1, and once in each of Sessions 2 and 3. The items in Session 1 were randomised such that: (1) each presentation of an item was separated by at least 20 interleaving items; and (2) no item was immediately preceded by the presentation of a phonologically, semantically, or associatively related item. The items in Sessions 2 and 3 were randomised following (2), but we additionally ensured that no more than five items of the same item type (new or old) were presented in adjacent trials.

*Procedure*

All sessions were administered online using Frinex (FRamework for INteractive EXperiments; a software package developed for running experiments by the technical group at the Max Planck Institute for Psycholinguistics). Although research suggests that collecting speech production data online may be noisier than in the laboratory, with longer tails in the response time distribution, onset latencies can be measured with good accuracy (e.g., Fairs & Strijkers, 2021; Stark, van Scherpenberg, Obrig, & Abdel Rahman, 2023; Vogt, Hauber, Kuhlen, & Abdel Rahman, 2021). Importantly for our purposes, these studies have demonstrated that the word frequency effect can be elicited in online experiments. For example, Fairs and Strijkers found a word frequency effect both online and in the laboratory. Although latencies collected online were longer, there was no significant difference in the size of the word frequency effect in the two experiments.

Participants were encouraged to complete the experiment in a quiet environment, away from any distractions such as phones or televisions. Each trial began with a fixation cross (+) presented in the centre of the screen for 500 ms. The fixation cross then disappeared, and the target picture was presented 300 ms later in the centre of the screen.

Participants had 2000 ms to name the picture before it disappeared and the next trial began automatically. If they named the picture before the end of the 2000 ms, they could click a "Volgende" ("Next") button on-screen to begin the next trial. The next trial began 1500 ms later, either after the timeout or after participants pressed the button. Each picture was preloaded at the start of the experiment to ensure there were no delays in image presentation once the trial started.

Before beginning each session, participants checked their microphone was working by creating a test recording and listening to the audio playback to ensure they could be clearly heard. Participants completed four practice trials to familiarise themselves with the experimental procedure, and then began the main experiment. For Session 1, the 384 pictures (six presentations of each of the 64 items) were divided into four blocks of 96 trials, and participants were given the opportunity to take a break after each block. For Sessions 2 and 3, each picture was presented once and the 128 pictures were divided into two blocks of 64 trials.

Participants completed Session 2 between three and five nights after Session 1 ($M = 4$ nights) and Session 3 between ten and fourteen nights after Session 2 ($M = 12$ nights). There were two participants who completed Session 2 six nights after Session 1 and three participants who completed Session 3 more than 14 nights after Session 2 (two completed after 15 nights; one after 16 nights). We included these participants in the analysis and report these models, but we also ran an analysis without their data and the results still held.

**Data analysis**

Picture naming latencies were measured from picture onset and were manually measured in Praat by trained Dutch speakers. Before analysis, we discarded 357 trials (2.21%) from Session 1, 273 trials (5.61%) from Session 2, and 177 trials (3.55%) from

Session 3 because participants either did not provide an answer within the 2000 ms time limit, or because the audio file was corrupt and we could not determine what the participant said. We also discarded 1525 trials (9.46%) from Session 1, 818 trials (16.82%) from Session 2, and 765 trials (15.21%) from Session 3 because participants named the picture other than expected, produced a nonspeech sound, a disfluency, or an utterance repair. This left us with 14246 trials for analysis for Session 1, 3773 for Session 2, and 4050 for Session 3. A breakdown of exclusions for high and low frequency words on each presentation can be found in Appendix B. These exclusion rates are high in comparison to laboratory studies, but they are similar to exclusion rates reported in other online studies (e.g., Fairs & Strijkers, 2021).

For Session 1, we evaluated the effects of frequency and presentation on response latencies using linear mixed effects models (Baayen, Davidson, & Bates, 2008) using the *lmer* function of the *lme4* package (version 1.1-26; Bates, Maechler, Bolker, & Walker, 2021) in RStudio (version 1.2.5042). Response latencies were predicted by Frequency (reference level: low vs. high), which was contrast coded (-0.5, 0.5). We also included Presentation and its interaction with Frequency in the analysis. We initially fitted a model using the maximal random effects structure justified by our design (Barr, Levy, Scheepers, & Tily, 2013), with correlations among random effects set to zero to aid model convergence. However, this model failed to converge, and so the final model included by-item intercepts and by-participant random effects for Presentation, Frequency and their interaction. Full model formulas and outputs are reported in Appendix A.

We fitted similar models to Sessions 2 and 3. Response latencies were again predicted by Frequency, but we also included Item Type (reference level: new vs. old) and its interaction with Frequency in the analysis. Item Type was contrast coded (-0.5, 0.5). For Session 2, we did not include the interaction between Frequency and Item Type as a by-participant random effect because doing so resulted in a singular fit error. For Session 3, we used the maximal random effects structure. Correlations among random effects were again set to zero to aid model convergence.

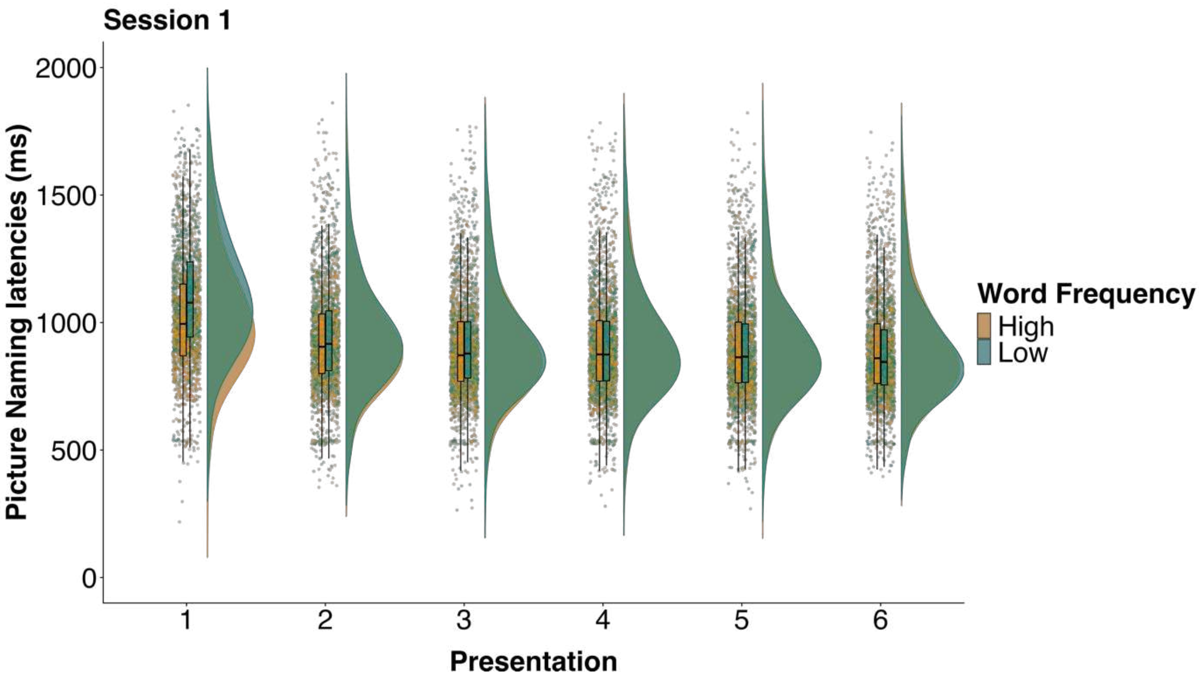
We report coefficient estimates (*b*), standard errors (*SE*), and *t*-values for each predictor. We assume that a *t*-value of $\pm$ 1.96 or greater indicates significance at the 0.05 alpha level (Baayen et al., 2008). The raw data and analysis scripts are available at: https://osf.io/pa8tb/.

## Results

### *Session 1*

On average, participants responded 929 ms (Figure 1) after picture onset. Participants were faster to name pictures with HF ($M$ = 928 ms) than LF names ($M$ = 940 ms; $b$ = -68.11, $SE$ = 12.78, $t$ = -5.33). Participants' naming latencies decreased with each presentation (Presentation 1 $M$ = 1064 ms; Presentation 2 $M$ = 942 ms; Presentation 3 $M$ = 910 ms; Presentation 4 $M$ = 909 ms; Presentation 5 $M$ = 902 ms; Presentation 6 $M$ = 893 ms; $b$ = -27.93, $SE$ = 2.98; $t$ = -9.36).

Figure 1. Distribution (and individual data points) of naming latencies for high and low frequency pictures on each presentation.



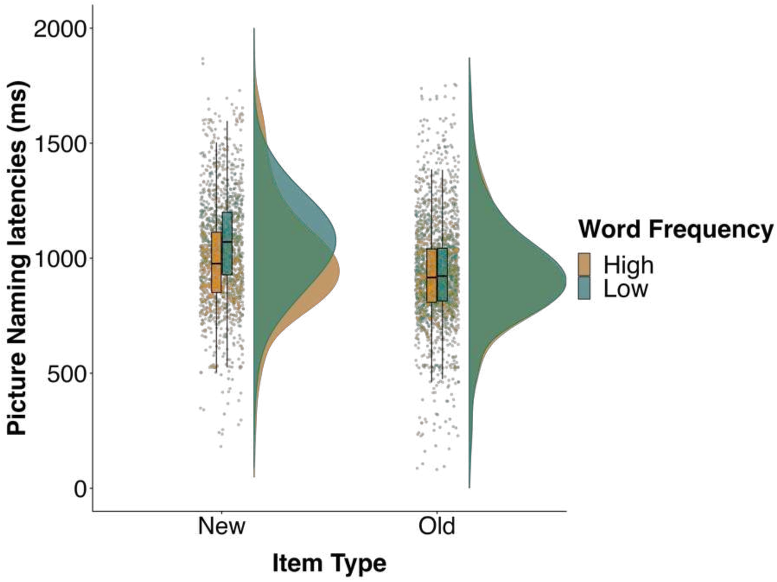

We also found a significant interaction between Frequency and Presentation (*b* = 15.37, *SE* = 1.50, *t* = 10.24). We followed up this interaction using the emmeans package (version 1.8.7) to compute simple pairwise comparisons for an effect of Frequency at each Presentation level with Bonferroni corrections. The model subtracted the LF effect from the HF effect on each presentation, and so returned positive estimates when participants were faster to name HF than LF pictures and negative estimates when they were faster to name LF pictures. Participants were faster to name pictures with HF rather than LF names on their first (*b* = 52.73, *SE* = 12.30, $p < .001$) and second presentation (*b* = 37.36, *SE* = 11.90, *p* = .002), but not on their third (*b* = 21.99, *SE* = 11.70, *p* = .06), fourth (*b* = 6.62, *SE* = 11.80, *p* = .57), or fifth presentation (*b* = -8.76, *SE* = 12.00, *p* = 0.47). There was some evidence for a Frequency effect on the sixth presentation (*b* = -24.13, *SE* = 12.40, *p* = .05), but this effect was only marginally significant and in the opposite direction of what we expected (i.e., faster for LF than HF pictures), so we do not interpret it further. Thus, we replicated previous studies (Corps & Meyer, 2023) and found that the word frequency effect was eliminated by repeated naming.[1]

*Session 2*

On average, participants responded 973 ms (Figure 2) after picture onset. They were faster to name pictures with HF (*M* = 971 ms) than LF (*M* = 992 ms; *b* = -41.28, *SE* = 15.07, *t* = -2.74). Participants were also faster to name old items (*M* = 941 ms), which they had named in Session 1, than new items, which they had not named before (*M* = 1031 ms; *b* = -113.78, *SE* = 15.48, *t* = -7.35).

[1] Note that the interaction between Frequency and Presentation could have occurred because the frequency effect is largest when response latencies are slowest (i.e., when naming is most difficult). However, this possibility seems unlikely given that we observed the same pattern of results when we log-transformed response latencies (see https://osf.io/pa8tb/).

Figure 2. Distribution (and individual data points) of naming latencies for high and low frequency pictures for old (named in Session 1) and new items (not named in Session 1) in Session 2.




We also found a significant interaction between Frequency and Item Type ($b$ = 66.64, $SE$ = 30.12, $t$ = 2.21). We followed up this interaction using the same procedure as Session 1. Participants were faster to name pictures with HF rather than LF names when these pictures were new ($b$ = 74.60, $SE$ = 21.80, $p$ < .001) but not when they were old ($b$ = 7.96, $SE$ = 20.80, $p$ = .70). Thus, the word frequency effect for the old items, which had been present in the first and second presentation of Session 1 but disappeared in the following presentations, was still absent three to five nights later in Session 2. This finding suggests that naming the pictures in Session 1 had a durable effect on the accessibility of the object names.

The word frequency effect for the new items, but not the old items, supports the conclusion that the frequency effect for old items was eliminated through their use in Session 1. The validity of the comparison of the results for old and new items rests on the assumption that two sets are comparable in terms of the relevant conceptual and lexical properties. Table 1 shows that they were well matched in terms of word frequency, name agreement, and age of acquisition. However, we discarded a larger portion of trials for the new items in Sessions 2

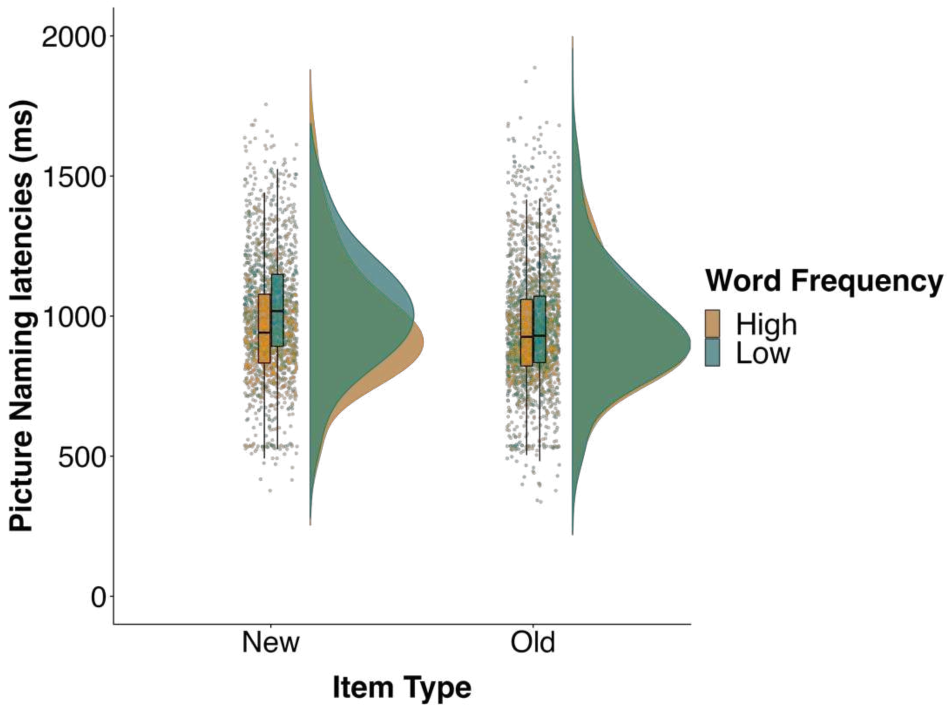
(799 trials; 32.85%) and 3 (614 trials; 24.60%) than on the first presentation of the pictures in Session 1 (515 trials; 19.16%), suggesting that the new items may have been more difficult to name than the old ones. To further strengthen our conclusion, we compared naming latencies for new items in Session 2 to naming latencies on Presentation 1 in Session 1 (i.e., instances where the two sets were both "new" to participants). Response latencies were again predicted by Frequency, but we also included Session (reference level: 2 vs. 1) and its interaction with Frequency in the analysis. Session was contrast coded (-0.5, 0.5). The model using the maximal random effects structure produced a singular fit error, likely because Frequency explained little by-participant variance. A model including by-participant random effects for Session and the interaction between Frequency and Session failed to converge, and so the final random effects structure included by-participant random effects for Session only. Correlations among random effects were again set to zero to aid model convergence.

Participants were again faster to name pictures with HF ($M$ = 1016 ms) rather than LF names ($M$ = 1097 ms; $b$ = -80.15, $SE$ = 18.53, $t$ = -4.33). Importantly, there was no difference in naming latencies for new pictures in Session 2 ($M$ = 1031 ms) and those presented for the first time in Session 1 ($M$ =1064 ms; $b$ = -22.07, $SE$ = 40.48, $t$ = -0.55), nor was there an interaction between Frequency and Session ($b$ = 8.21, $SE$ = 37.05, $t$ = 0.22). As a result, the presence of a frequency effect for new, but not old, items in Session 2 cannot be attributed to differences in item properties, and instead occurred because only the old, but not the new items had been named previously.

*Session 3*

On average, participants responded 973 ms (Figure 3) after picture onset. As in Sessions 1 and 2, participants were faster to name pictures with HF (*M* = 971 ms) than LF names (*M* = 996 ms; *b* = -35.86, *SE* = 13.49, *t* = -2.66). Participants were again faster to name old items (*M* = 964 ms) than new items (*M* = 1005 ms; *b* = -58.52, *SE* = 13.23, *t* = -4.22).

Figure 3. Distribution (and individual data points) of picture naming latencies for high and low frequency pictures for old (named in Session 1) and new items (not named in Session 1) in Session 3.




We found a marginally significant interaction between Frequency and Item Type (*b* = 51.57, *SE* = 26.49, *t* = 1.95), and we followed up this interaction using the same procedure as Session 2. As in Session 2, participants were faster to name pictures with HF than LF names when these pictures were new (*b* = 61.70, *SE* = 19.20, *p* = .001), but not when they were old (*b* = 10.10, *SE* = 18.60, *p* = .59). Thus, the frequency effect that was absent in Session 2 was still absent ten to fourteen nights later in Session 3.

*Comparing responses to new items in Sessions 2 and 3*

Participants named the same new items (i.e., not seen in Session 1) in Sessions 2 and 3, giving us the opportunity to test whether repeatedly naming the same picture across sessions reduced the word frequency effect. To do so, we fitted a model in which naming latencies to the new items in both sessions were predicted by Frequency (reference level: low vs. high), Session (reference level: 2 vs. 3), and their interaction. Both predictors were contrast coded (-0.5, 0.5). We included by-item random effects for Session and by-participant random effects for Frequency because including Session and its interaction with Frequency produced a singular fit error. Correlations among random effects were fixed to zero to aid model convergence.

Much like observed for the old pictures, participants were faster to name new pictures with HF ($M$ = 994 ms) rather than LF names ($M$ = 1058 ms; $b$ = -70.80, $SE$ = 23.97, $t$ = -2.95). However, there were no differences in naming latencies for the new pictures between Sessions 2 ($M$ = 1031 ms) and 3 ($M$ = 1005 ms; $b$ = 34.55, $SE$ = 37.72, $t$ = 0.92) and there was no interaction between Frequency and Session ($b$ = -19.50, $SE$ = 13.39, $t$ = -1.46). Thus, naming new pictures once in Session 2 was not enough to reduce the word frequency effect when they were named again in Session 3.

**Discussion**

In Experiment 1, we tested whether producing well-known but infrequent words affected their long-term accessibility. Participants named HF and LF pictures six times in a single session and then returned and named the same HF and LF pictures, together with new HF and LF pictures, in a second and third session. We found a word frequency effect the first and second time participants named the pictures, but not the subsequent four times. Thus, consistent with previous research (e.g., Corps & Meyer, 2023), the word frequency effect disappeared with repeated naming. Importantly, the word frequency effect was still absent in both Session 2 (three to five nights after Session 1) and 3 (10 to 14 nights after Session 2). Thus, production of well-known but infrequent words affected their long-term accessibility, suggesting that the word frequency effect arises as a result of accumulating experience with words.

It is worth noting that Sessions 2 and 3 showed similar results for the old items, with identical average naming latencies in the two sessions. This finding could suggest that any new learning in Session 2 had minimal impact on naming times in Session 3. Alternatively, it could be that naming in Session 2 refreshed participants' memory for the pictures in Session 3. But regardless, the interval between Sessions 2 and 3 was long (10 to 14 nights), and so even if Session 2 refreshed participants' memory, the findings from Session 3 still suggest that producing a word can affect its long-term accessibility.

For the new items, that is the items not presented in Session 1, we found a frequency effect in both Session 2 and Session 3. In other words, naming these items once in Session 2 was not enough to reduce the word frequency effect when they were named again in Session 3. This finding makes sense, given that the frequency effect in Session 1 was only eliminated after two presentations of the pictures, but the participants only named the new items once in Session 2. We return to this pattern in the General Discussion.

Thus, Experiment 1 suggests that producing words multiple times in short succession alters their long-term accessibility, reducing the word frequency effect. However, it is possible that at least part of the repetition effect arose during visual and conceptual processing, rather than at the lexical level, given that research suggests picture recognition strongly benefits from repetition (e.g., Lachman et al., 1980). We addressed this issue in Experiment 2, in which participants produced words without their accompanying pictures in Session 1. After naming each word, participants judged whether a second word indicated a property of the concept denoted by the word that they had just named. Participants named and made judgements about each word six times, as in Experiment 1. Thus, Session 1 served to expose participants to the picture names, elicit their pronunciations, and encourage activation of some aspects of the word's meaning and its semantic features in the same way as in Experiment 1, but without exposing participants to the pictures. In Session 2, participants then named old HF and LF pictures, whose names they had seen and produced in Session 1, together with new HF and LF pictures, whose names they had not seen and produced. Thus, Experiment 2 separated effects of picture repetition from word repetition. Since Sessions 2 and 3 showed similar results in Experiment 1, participants in Experiment 2 completed a second session only, which was administered seven to ten days after Session 1. As in Experiment 1, we compared the word frequency effect for the old and new items in Session 2.

## Experiment 2

### Method

#### *Participants*

Forty-three native Dutch speakers (32 females, 11 males; *M*age = 32 years) were recruited using the same procedure as Experiment 1 and participated in exchange for 16 euros. We discarded data from two participants for Session 1 due to technical errors, and so this analysis is based on data from 41 participants (32 females, 9 males; *M*age = 32 years). Three participants dropped out after Session 1, while two participants provided incomprehensible audio recordings, and so the analysis for this session is based on the data from 38 participants (28 females, 10 males; *M*age = 32 years).

*Materials and procedure*

For Session 1, we used the 64 picture names from Experiment 1 but paired each one with six properties, which were selected and validated by four native Dutch speakers. Four of these properties were related to the picture name, while two were unrelated. For example, the picture name *aap* (monkey) was paired with the four correct properties *dier* (animal), *dierentuin* (zoo), *levend* (living), and *lenig* (limber) and the two incorrect properties *gebouw* (building) and *paars* (purple; see OSF for a full list). Each property was paired with no more than two picture names. Properties were between one and five syllables in length ($M$ = 2.02) and had SUBTLEX frequencies ranging from one to 49718 ($M$ = 1969).

Each trial began with a fixation cross (+) presented in the centre of the screen for 500 ms. The fixation cross then disappeared, and the target word was presented 300 ms later in the centre of the screen in lowercase black Courier 32-point typeface. Participants had 2000 ms to produce the word before it disappeared. A second word was then presented 500 ms later in the same format as the first word. Participants responded *yes* (M key on their keyboard) if the second word was a property of the first, or *no* (Z key on their keyboard) if it was not. The next trial began 1500 ms after participants responded. Each word was paired with four correct and two incorrect properties, and so the probability of a *yes* response was 60%. Thus, each word was named six times.

Session 2 used the same stimuli and procedure as Experiment 1. Participants completed Session 2 between seven and ten nights after Session 1 ($M$ = 9 nights). There was one participant who completed Session 2 six nights after Session 1 and four participants who completed Session 2 more than ten nights after Session 1 (three completed after 11 nights; one after 12). We included these participants in the analysis, but we also ran an analysis without their data to ensure the findings still held.

The stimuli in both sessions were randomised using the same procedure as Experiment 1. Participants checked their microphone was working by creating a test recording and listening to the audio playback to ensure they could be heard clearly. They then completed four practice trials to familiarise themselves with the experimental procedure, before beginning the main experiment.

**Data Analysis**

Word and picture naming latencies were calculated using the same procedure as Experiment 1. Before analysis, we discarded 126 trials (0.80%) from Session 1 and 251 trials (5.16%) from Session 2 because participants either did not provide an answer within the 2000 ms time limit, or because the audio file was corrupt, and we could not determine what the participant said. We also discarded 27 trials (0.17%) from Session 1 and 874 trials (17.97%) from Session 2 because participants named the word or picture other than expected, produced a nonspeech sound, a disfluency, or an utterance repair. This left us with 15591 trials for analysis for Session 1 and 3739 trials for Session 2.

For Session 1, we focused our analysis on the word naming latencies. It was not clear whether decision latencies in the property judgement task would be affected by the frequency of the first word (that participants named), the second word (that was a property of the first), or both. Furthermore, the primary function of the property judgement trials was to encourage participants to repeatedly activate the meaning of the word they previously named without creating a link between a picture and a particular name. We do, however, report exploratory analysis of the property judgements in Appendix C. For Session 2, we focused our analysis on the picture naming latencies.

We analysed the data using the same procedure as for Experiment 1. The maximal model produced a singular fit error for Session 1, likely because the by-participant effect for

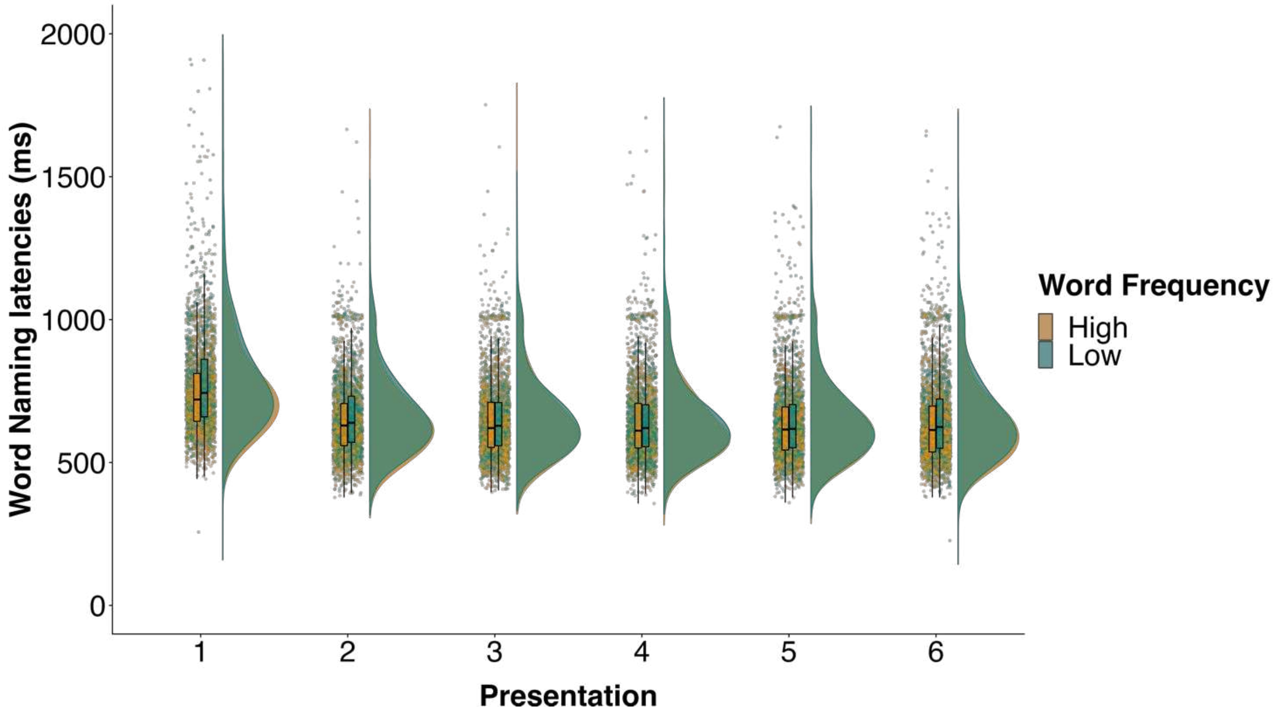
Frequency explained little variance. As a result, the final model included by-participant random effects for Frequency and Presentation and by-item random effects for Presentation. For Session 2, the maximal model produced a singular fit error because the by-participant random effects for Item Type and its interaction with Frequency explained little variance. As a result, the final model included by-participant random effects for Frequency only. In all analyses, correlations among random effects were fixed to zero to aid model convergence.

**Results and Discussion**

*Session 1*

On average, participants responded 668 ms (Figure 4) after word onset. As in Experiment 1, participants were faster to name HF (*M* = 662 ms) than LF words (*M* = 675 ms; *b* = -24.14, *SE* = 11.33, *t* = -2.13) and their naming latencies decreased with each presentation (Presentation 1 *M* = 766 ms; Presentation 2 *M* = 657 ms; Presentation 3 *M* = 651 ms; Presentation 4 *M* = 646 ms; Presentation 5 *M* = 644 ms; Presentation 6 *M* = 649 ms; *b* = -17.72, *SE* = 2.54, *t* = -6.99). Consistent with previous research (e.g., Wheeldon & Monsell, 1992), participants were faster when naming words in Experiment 2 compared to pictures in Experiment 1, and the repetition priming effect was smaller for words (a coefficient of -17.72) than for pictures (a coefficient of -27.93).

Figure 4. Distribution (and individual data points) of word naming latencies for high and low frequency words on each presentation.




Unlike Experiment 1, we found no interaction between Frequency and Presentation (*b* = 3.28, *SE* = 2.03, *t* = 1.61). This finding could suggest that the interaction in Experiment 1 occurred partly because repetition benefited picture recognition, or there may have been less room for an interaction to occur because the word frequency effect was relatively small (13 ms). Although the interaction was not significant, we still followed up the interaction using the emmeans package for consistency with Experiment 1 and to investigate whether there was any evidence that the frequency effect was larger on the first and second presentation than on the other four. Participants were marginally faster to name HF than LF words on their first presentation (*b* = 20.86, *SE* = 11.20, *p* = .06), but not on their second (*b* = 17.57, *SE* =

11.40, $p = .12$), third ($b = 14.29$, $SE = 11.90$, $p = .23$), fourth ($b = 11.01$, $SE = 12.80$, $p = .39$), fifth ($b = 7.73$, $SE = 13.90$, $p = .58$), or sixth presentation ($b = 4.44$, $SE = 15.20$, $p = .77$).[2]

Thus, there was some evidence that repeated naming and property judgement reduced the word frequency effect after a single presentation of the items, as in Experiment 1. The word frequency effect was smaller on the item's first presentation in Experiment 2 (a coefficient of 20.86), where participants could name the word by simply retrieving its form, than in Experiment 1 (a coefficient of 52.73), where participants had to retrieve the picture's concept and its form, consistent with studies that suggest word frequency is not just located at the word form level (e.g., Corps & Meyer, 2023; Kittredge et al., 2008).

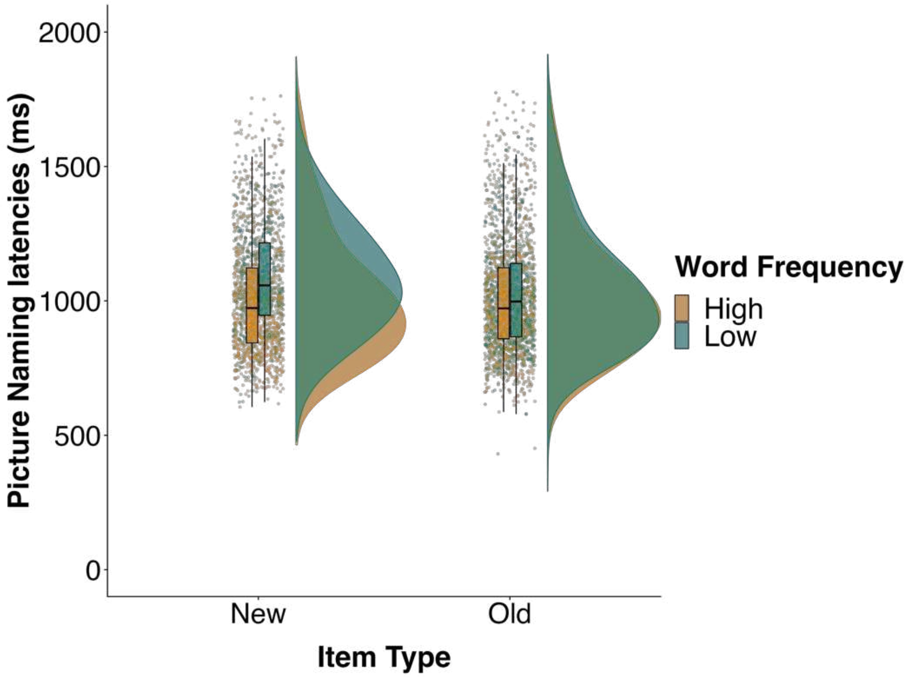
The robust repetition effect and the word frequency effect seen on the first presentation of the materials replicate key findings in the literature on visual word processing using the word naming or lexical decision tasks (e.g., Balota & Spieler, 1999; MacLeod & Masson, 2000; Monsell, Doyle & Haggard, 1989; Murray & Forster, 2004). Studies investigating the interaction between word frequency and repetition effects have reported mixed results. In some studies, often using masked primes and the lexical decision task, the two effects were additive (e.g., Grainger et al., 2012; Kinoshita, 2006), whereas in studies using overt primes and the naming task, they typically interacted, as was the case in the present study (e.g., Kamienkowski et al, 2018; Lowder, Choi, & Gordon, 2013; but see Murray & Forster, 2004).

[2] As in Experiment 1, this pattern could have occurred because the frequency effect is largest when response latencies are slowest (i.e., when naming is most difficult). Again, this possibility seems unlikely given that we observed the same pattern of results when we log-transformed response latencies.

*Session 2*

On average, participants responded 1028 ms (Figure 5) after picture onset. They were faster to name pictures with HF (*M* = 1013 ms) than LF names (*M* = 1061 ms; *b* = -55.32, *SE* = 18.28, *t* = -3.03). As in Experiment 1, participants were faster to name old (*M* = 1025 ms) than new items (*M* = 1045 ms; *b* = 41.94, *SE* = 17.98, *t* = 2.33). The difference between new and old items was smaller in this experiment (19 ms) than in Experiment 1 (90 ms). This difference likely occurred because the participants of Experiment 1 had seen the target pictures in Session 1, whereas the participants of Experiment 2 had only read and produced their names. Thus, part of the large difference in naming latencies for old and new items in Experiment 1 was likely to due to repetition priming arising during the visual and conceptual processing of the pictures (e.g., Francis, Corral, Jones, & Sáenz, 2008; Francis & Sáenz, 2007).

Figure 5. Distribution (and individual data points) of picture naming latencies for high and low pictures for old (picture name produced in Session 1) and new items (picture name not produced in Session 1) in Session 2.




We found a significant interaction between Frequency and Item Type ($b$ = -73.77, $SE$ = 35.97, $t$ = -2.05) and we followed up this interaction using the same procedure as Experiment 1. As in Experiment 1, participants were faster to name pictures with HF than LF names when these pictures were new ($b$ = 92.20, $SE$ = 26.0, $p$ < .001) but not when they were old ($b$ = 18.40, $SE$ = 25.30, $p$ = .47).

In sum, the word frequency effect was reduced after a single presentation in Session 1 and was still absent in Session 2, even when participants had not seen the pictures in Session 1. Thus, the effect cannot be related to familiarity with or recognition of the pictures and was instead related to recent use of the picture names. Consistent with Experiment 1, these findings suggest that repeatedly using a word alters its accessibility, making it easier to produce later.

## General Discussion

In two experiments, we investigated whether producing well-known but infrequent words affects their long-term accessibility. In both experiments, participants were faster to name HF than LF pictures (Experiment 1) and words (Experiment 2) but this word frequency effect was eliminated with multiple presentations of the items. This word frequency effect was still absent up to two weeks later when participants returned and named old HF and LF pictures, whose names they had produced in Session 1, together with new HF and LF pictures, whose names they had not previously produced. Together, these findings suggest that producing well-known but infrequent words affects their long-term accessibility, such that an existing word frequency effect is overridden. Thus, our results suggest that using a word leads to memory traces that are durable enough to support cumulative learning, and the word frequency effect arises as a result of accumulating experience with words. This assumption is standard in theories of lexical access (e.g., Levelt, Roelofs, & Meyer, 1999) but has rarely been tested directly.

Our findings fit with the broad class of adaptive models of lexical access (e.g., Damian & Als, 2005; Howard et al., 2006; Oppenheim et al., 2010), which assume that the structure of the lexical network is altered through retrieval processes. In connectionist models, using a word strengthens the connections from the relevant semantic features to the lexical representations of selected target words (increasing future accessibility) and weakens the links to any unselected competitors (decreasing future accessibility; e.g., Dell, Chang, & Griffin, 1999). For example, when naming a picture of a dog participants do not only activate the word *dog*, but also activate related words such as *cat, rabbit,* and *doll*. When *dog* is selected to be produced, the connections between the active semantic features and the lexical representation for *dog* are strengthened, while the connections to the unselected words *cat, rabbit,* and *doll* are weakened. As a result, *dog* becomes more easily accessible. Adaptive models of lexical access typically assume error-based learning mechanisms, such that strengthening of links has a stronger impact on relatively weak links (e.g., links to infrequent or unexpected words) than on stronger links. Such error-based learning can explain why the word frequency effect is reduced or eliminated when words are presented or produced several times within a session: LF words benefit from the weight changes more than HF words.

We observed a durable between-session repetition priming effect only when participants had previously produced the picture name six times (from Session 1 to 2), but not when they had produced it only once (from Session 2 to 3 in Experiment 1). This finding suggests that the number of retrieval episodes within a session affected how much the word's memory representations were changed, which in turn affected the long-term durability of these weight changes, consistent with error-based learning accounts (e.g., Oppenheim et al., 2010). Weight changes will decay over time, but when a picture is named repeatedly within a short time span, the decay is counteracted and the connections between semantic features and lexical representations are further strengthened. Our findings suggest that such cumulative learning is necessary for longer-term repetition priming and the elimination of the word frequency effect. On the basis of our results, we cannot determine how often a word needs to be produced for longer-term repetition priming to arise. Based on the comparison of the results obtained for the old items, which were named six times in session 1, and the new items, which were only named once, we can only conclude that six naming episodes appear to be sufficient, whereas one naming episode is not. Our results are consistent with the memory literature, which has shown that repetition of materials and overlearning improve memory (e.g., Hintzman, 1976; Roediger & Butler, 2011; Underwood, 1969), and the repetition priming literature, which has shown that multiple primes lead to stronger effects than single primes (e.g., Betts et al., 2018; Durso & Johnson, 1979; Forster & Davis, 1984).

A controversial issue is which learning mechanisms support longer-term lexical learning. Gaskell, Cairney, and Rodd (2019; see also Rodd et al., 2013, 2016) suggest that repetition priming could have long-term effects in two different ways. Under their *immediate alteration* account, each experience with a word immediately alters the long-term connections between a word's semantic features and its lexical representations. Thus, long-term representations are altered directly. In contrast, their *contextual binding* account suggests that hippocampal resources are recruited on each new experience of a word, temporarily binding the word and its context to provide an additional source of knowledge alongside permanent lexical knowledge. These hippocampal representations are then integrated into the long-term lexical network after a period of consolidation. Consolidation primarily occurs during sleep and is associated with the occurrence of sleep spindles (e.g., Tamminem et al., 2010). Contextual binding, followed by consolidation, has primarily been discussed in relation to novel word and grammar learning, but Gaskell et al. proposed that it also applies to word meaning modification when listeners are exposed to the less frequent meaning of a homophone (see also Curtis et al., 2022; Mak, Curtis, Rodd, & Gaskell, 2023). Extrapolating from these findings, one might expect that durable changes in a word's accessibility after use also require a process of consolidation. Further work is needed to assess this hypothesis, but our results inform this work and theories about the plasticity of the adult mental lexicon by demonstrating that recent use of words has both an immediate and long-term impact on accessibility. We observed long-term effects only when participants produced words several times within a session, suggesting multiple repetitions are necessary for long-term changes to the mental lexicon.

The key finding of our work is that the frequency effect was reduced for old items in Sessions 2 and 3, pointing to durable changes in participants' lexical representations. But we also observed that the word frequency effect was eliminated by repetition within a single session, regardless of whether participants named pictures or words. This finding replicates our previous picture naming study (Corps & Meyer, 2023), in which we found a strong frequency effect when participants named pictures the first time, but not when they named the pictures a second or third time. We also found a similar pattern of results when participants repeatedly judged the gender of the picture names. The similarity of the results in the two tasks suggest that they tap into similar representations that are sensitive to word frequency and repetition. These representations could be located at the lexical level, the word form level, or both (Kittredge et al., 2008). Further research could investigate whether the repetition effects observed here also occur in tasks that isolate these individual levels (such as isolated word naming or lexical decision), further determining the locus of the word frequency effect. Regardless, our findings suggest that repetition eliminates the word frequency effect.

However, other picture naming studies have shown that the word frequency effect was substantially reduced by repetition but still present (e.g., Griffin & Bock, 1998) and Jescheniak and Levelt (1994) even found that the effect was stable across multiple presentations of the same items. Differences in the strength of the interaction between word frequency and repetition across studies could arise for a number of reasons. One of them concerns the choice of materials. Studies have varied in the way differences in word frequency are implemented, and particularly in how much high and low frequency items differ from each other in frequency, and in whether very frequent or very infrequent items are included in the materials. In our study, the difference between the HF (41.56 per million words on average) and LF (1.23 per million words) conditions was smaller than in previous studies when considering absolute frequency (Griffin & Bock: HF = 110 per million, LF = 15 per million; Jescheniak & Levelt: HF = 150.70 per million, LF = 6 per million), but larger when frequency was estimated on a log-scale (i.e., Zipf; our study HF = 4.51; LF = 2.98;

Jescheniak & Levelt HF = 3.48, LF = 2.13). However, these differences cannot fully explain why the word frequency effect disappeared with repetition in our study but not in others: Corps and Meyer (2023) found that the word frequency effect was eliminated by repetition even when using Jescheniak and Levelt's original materials. The difference might instead be attributable to differences in statistical power. Griffin and Bock tested 30 participants with 60 pictures and three presentations, Jescheniak and Levelt tested 12 participants with 48 pictures and three repetitions, Levelt et al. (1998) tested eight participants with 24 pictures and twelve repetitions, and Corps and Meyer tested around 40 participants with 64 pictures and three repetitions, which gave the study more power to detect an interaction between Frequency and Presentation. Nonetheless, the characteristics of the picture names are bound to affect the size of the frequency effect and the way it is changed when pictures are named several times.

Another set of factors that may affect how strongly frequency effects change with repetition of the materials concerns the participants' task. We found an interaction between frequency and repetition in Session 1 when participants named pictures (Experiment 1) but not when they named words and made semantic judgements (Experiment 2). However, both tasks led to the long-term elimination of the word frequency effect in Session 2. There is a large literature suggesting that memory for words depends on the task at encoding. For example, generating picture names leads to better memory than reading words (Bertsch, Pesta, Wiscott, & McDaniel, 2007; Slamecka & Graf, 1978; Zormpa, Brehm, Hoedemaker, & Meyer, 2019), and reading words aloud leads to better memory for the words than reading them silently (e.g., Conway & Gathercole, 1987; Fawcett, Quinlan, & Taylor, 2012; MacLeod & Bodner, 2017). In light of these results, one might expect the strength and longevity of repetition effects to also depend on encoding modality. Our results show that repeatedly producing HF and LF words as picture names or in a reading task in conjunction with a semantic judgement task eliminates the word frequency effect. Further research is needed to investigate whether the same holds for repeated silent reading of or listening to words.

Likewise, it is important to study how word frequency effects are altered when assessed in tasks other than picture naming, such as naming written words or sentence or text reading. There is a large literature on the interaction of word frequency and repetition effects in these tasks, and many studies have found that the word frequency effect is attenuated when items are repeated within a single experimental session (e.g., Besner & Swan, 1982; Drieghe & Chan Seem, 2022; Kamienkowski et al., 2018). However, the durability of these interactive effects has not been systematically investigated. Our prediction, based on the current data, is that there should be a durable reduction of the word frequency effect regardless of the task as long as sufficient cumulative learning has taken place within a session. Assessing this issue is important because we encounter words in different modalities in everyday life, probably more often in reading and listening than in speaking. To understand how word frequency effects build up and change in individuals' mental lexica, it is necessary to determine how much and how durably their lexical network is altered when words are encountered in different modalities.

Word frequency is a measure of how often a speaker has encountered a word. It is related to other variables, such as age of acquisition (e.g., Morrison & Ellis, 1995), word prevalence (e.g., Brysbaert, Stevens, Mandera, & Keuleers, 2016), and recency of use (e.g., Scarborough, Cortese, & Scarborough, 1977). An interesting recent proposal is that much (or all) of the effects of word frequency can be attributed to contextual diversity, a highly correlated variable which refers to the number of different contexts a word has been encountered in. For example, Adelman, Brown, and Quesada (2006) found that word naming and lexical decision latencies were faster for words with high contextual diversity (encountered in more contexts) than those with lower diversity (encountered in fewer contexts). Importantly, word frequency effects were eliminated when contextual diversity was controlled, suggesting that word frequency effects may actually be effects of contextual diversity.

Theoretical accounts of contextual diversity effects often assume that the lexicon is structured in such a way that useful words (i.e., those used in many contexts) are most accessible (e.g., Anderson & Milson, 1989; Johns, Dye, & Jones, 2016). Under this account, one might expect that contextual diversity should interact with repetition, given that recent use of a word indicates that it is useful. In our materials, word frequency and contextual diversity (retrieved from SUBTLEX) were highly correlated ($r = .97$, $p < .001$) and so we were unable to test whether repetition affected contextual diversity independently of word frequency. These effects could be separated experimentally by manipulating an individual speaker's learning experience for specific words, such as by presenting words multiple times in the same or different contexts. This approach has been adopted by studies investigating novel word learning (e.g., Frances, Martin, & Duñabeitia, 2020; see also Raviv, Lupyan, & Green, 2022, for a review). Using a similar strategy, one could assess how the ease of access to known words is affected when they are encountered multiple times in the same or different contexts.

Our findings have important methodological implications. When researchers use word frequency as a tool to manipulate the difficulty of lexical access or speech planning, they should avoid repeatedly presenting the items. Such repeated exposure will reduce any word frequency effect. Alternatively, if variation in word frequency is a nuisance variable whose impact should be minimised, then researchers will benefit from repeatedly presenting the same items because it will encourage participants to respond similarly to LF and HF pictures.

In sum, the frequency of words in text corpora has been shown to systematically affect how fast and accurately individual speakers produce and understand words. This finding raises the question of how word frequency – a corpus variable – makes its way into an individual's lexicon. A highly plausible, yet rarely assessed, proposal is that words that appear frequently in corpora are also experienced frequently by individuals, and that they learn from each encounter with these words. We found support for this view in two naming studies. In particular, we found that participants were faster to name HF than LF pictures (Experiment 1) and words (Experiment 2), but this effect was eliminated by repetition. This word frequency effect was still absent up to two weeks later, suggesting that producing a word multiple times within a short time period alters its long-term accessibility and leads to the elimination of the word frequency effect.

## References

Adelman, J. S., Brown, G. D., & Quesada, J. F. (2006). Contextual diversity, not word frequency, determines word-naming and lexical decision times. *Psychological science*, *17*, 814-823.

Anderson, J. R., & Milson, R. (1989). Human memory: An adaptive perspective. *Psychological Review*, *96*, 703-719.

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of memory and language*, *59*, 390-412.

Balota, D. A., & Spieler, D. H. (1999). Word frequency, repetition, and lexicality effects in word recognition tasks: beyond measures of central tendency. *Journal of Experimental Psychology: General*, *128*, 32-55.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of memory and language*, *68*, 255-278.

Bates, D. M., Maechler, M., Bolker, B., & Walker, S. (2021). Lme4: Linear mixed-effects models using "Eigein" and S4 (R package version 1.1-26). Retrieved from http://CRAN.R-project.org/package=lme4.

Bertsch, S., Pesta, B. J., Wiscott, R., & McDaniel, M. A. (2007). The generation effect: A meta-analytic review. *Memory & cognition*, *35*, 201-210.

Besner, D. & Swan, M. (1982). Models of lexical access in visual word recognition. *The Quarterly Journal of Experimental Psychology, 34,* 313-325.

Betts, H. N., Gilbert, R. A., Cai, Z. G., Okedara, Z. B., & Rodd, J. M. (2018). Retuning of lexical-semantic representations: Repetition and spacing effects in word-meaning priming. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *44*, 1130-1150.

Brysbaert, M., Mandera, P., & Keuleers, E. (2018). The word frequency effect in word processing: An updated review. *Current directions in psychological science*, *27*, 45-50.

Brysbaert, M., Stevens, M., Mandera, P., & Keuleers, E. (2016). The impact of word prevalence on lexical decision times: Evidence from the Dutch Lexicon Project 2. *Journal of Experimental Psychology: Human Perception and Performance*, *42*, 441-458.

Conway, M. A., & Gathercole, S. E. (1987). Modality and long-term memory. *Journal of Memory and language*, *26*, 341-361.

Corps, R. E. (2015, February 13). Multiple repetitions lead to the long-term elimination of the word frequency effect. https://doi.org/10.17605/OSF.IO/PA8TB.

Corps, R. E., & Meyer, A. S. (2023). Word frequency has similar effects in picture naming and gender decision: A failure to replicate Jescheniak and Levelt (1994). *Acta Psychologica*, *241*, https://doi.org/10.1016/j.actpsy.2023.104073.

Curtis, A. J., Mak, M. H., Chen, S., Rodd, J. M., & Gaskell, M. G. (2022). Word-meaning priming extends beyond homonyms. *Cognition*, *226*, https://doi.org/10.1016/j.cognition.2022.105175.

Damian, M. F., & Als, L. C. (2005). Long-lasting semantic context effects in the spoken production of object names. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*, 1372-1384.

Decuyper, C., Brysbaert, M., Brodeur, M. B., & Meyer, A. S. (2021). Bank of Standardized Stimuli (BOSS): Dutch names for 1400 photographs. *Journal of Cognition*, *4*, https://doi.org/10.5334/joc.180.

Dell, G. S., Chang, F., & Griffin, Z. M. (1999). Connectionist models of language production: Lexical access and grammatical encoding. *Cognitive Science*, *23*, 517-542.

Drieghe, D. & Chan Seem, R. (2022). Parafoveal processing of repeated words during reading. *Psychonomic Bulletin & Review, 29,* 1451-1460.

Durso, F. T., & Johnson, M. K. (1979). Facilitation in naming and categorizing repeated pictures and words. *Journal of Experimental Psychology: Human Learning and Memory*, *5*, 449-459.

Estes, W. K., & Maddox, W. T. (2002). On the processes underlying stimulus-familiarity effects in recognition of words and nonwords. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*, 1003-1018.

Fairs, A., & Strijkers, K. (2021). Can we use the internet to study speech production? Yes we can! Evidence contrasting online versus laboratory naming latencies and errors. *Plos one*, *16*, https://doi.org/10.1371/journal. pone.0258908.

Fawcett, J. M., Quinlan, C. K., & Taylor, T. L. (2012). Interplay of the production and picture superiority effects: A signal detection analysis. *Memory*, *20*, 655-666.

Forster, K. I., & Davis, C. (1984). Repetition priming and frequency attenuation in lexical access. *Journal of experimental psychology: Learning, Memory, and Cognition*, *10*, 680-698.

Frances, C., Martin, C. D., & Duñabeitia, J. A. (2020). The effects of contextual diversity on incidental vocabulary learning in the native and a foreign language. *Scientific Reports, 10,* https://doi.org/10.1038/s41598-020-70922-1.

Francis, W. S. (2014). Repetition priming in picture naming: Sustained learning through the speeding of multiple processes. *Psychonomic bulletin & review*, *21*, 1301-1308.

Francis, W. S., Corral, N. I., Jones, M. L., & Sáenz, S. P. (2008). Decomposition of repetition priming components in picture naming. *Journal of Experimental Psychology: General*, *137*, 566-590.

Francis, W. S., Gurrola, B. V., & Martínez, M. (2022). Comprehension exposures to words in sentence contexts impact spoken word production. *Memory & Cognition*, *50*, 192-215.

Francis, W. S., & Sáenz, S. P. (2007). Repetition priming endurance in picture naming and translation: Contributions of component processes. *Memory & Cognition*, *35*, 481-493.

Gaskell, M. G., Cairney, S. A., & Rodd, J. M. (2019). Contextual priming of word meanings is stabilized over sleep. *Cognition*, *182*, 109-126.

Gollan, T. H., Montoya, R. I., Cera, C., & Sandoval, T. C. (2008). More use almost always means a smaller frequency effect: Aging, bilingualism, and the weaker links hypothesis. *Journal of Memory and Language*, *58*, 787-814.

Gollan, T. H., Montoya, R. I., Fennema-Notestine, C., & Morris, S. K. (2005). Bilingualism affects picture naming but not picture classification. *Memory & Cognition*, *33*, 1220-1234.

Grainger, J., Lopez, D., Eddy, M., Dufau, S., & Holcomb, P. J. (2012). How word frequency modulates masked repetition priming: An ERP investigation. *Psychophysiology*, *49*, 604-616.

Griffin, Z. M., & Bock, K. (1998). Constraint, word frequency, and the relationship between lexical processing levels in spoken word production. *Journal of Memory and Language*, *38*, 313-338.

Hintzman, D. L. (1976). Repetition and memory. *Psychology of learning and motivation*, *10*, 47-91.

Howard, D., Nickels, L., Coltheart, M., & Cole-Virtue, J. (2006). Cumulative semantic inhibition in picture naming: Experimental and computational studies. *Cognition*, *100*, 464-482.

Jescheniak, J. D., & Levelt, W. J. (1994). Word frequency effects in speech production: Retrieval of syntactic information and of phonological form. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*, 824-843.

Johns, B. T., Dye, M., & Jones, M. N. (2016). The influence of contextual diversity on word learning. *Psychonomic Bulletin & Review*, *23*, 1214-1220.

Johnson, C. J., Paivio, A., & Clark, J. M. (1996). Cognitive components of picture naming. *Psychological Bulletin*, *120*, 113-139.

Kamienkowski, J. E., Carbajal, M. J., Bianchi, B., Sigman, M., & Shalom, D. E. (2018). Cumulative repetition effects across multiple readings of a word: Evidence from eye movements. *Discourse Processes*, *55*, 256-271.

Kinoshita, S. (2006). Additive and interactive effects of word frequency and masked repetition in the lexical decision task. *Psychonomic Bulletin & Review*, *13*, 668-673.

Kittredge, A. K., Dell, G. S., Verkuilen, J., & Schwartz, M. F. (2008). Where is the effect of frequency in word production? Insights from aphasic picture-naming errors. *Cognitive Neuropsychology*, *25*, 463-492.

Lachman, R., Lachman, J. L., Thronesbery, C., & Sala, L. S. (1980). Object salience and code separation in picture naming. *Bulletin of the Psychonomic Society*, *16*, 187-190.

La Heij, W., Puerta-Melguizo, M. C., van Oostrum, M., & Starreveld, P. A. (1999). Picture naming: Identical priming and word frequency interact. *Acta psychologica*, *102*, 77-95.

Levelt, W. J., Praamstra, P., Meyer, A. S., Helenius, P., & Salmelin, R. (1998). An MEG study of picture naming. *Journal of Cognitive Neuroscience*, *10*, 553-567.

Levelt, W. J., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences*, *22*, 1-38.

Lowder, M. W., Choi, W., & Gordon, P. C. (2013). Word recognition during reading: The interaction between lexical repetition and frequency. *Memory & Cognition*, *41*, 738-751.

MacLeod, C. M., & Bodner, G. E. (2017). The production effect in memory. *Current Directions in Psychological Science*, *26*, 390-395.

MacLeod, C. M., & Masson, M. E. (2000). Repetition priming in speeded word reading: Contributions of perceptual and conceptual processing episodes. *Journal of Memory and Language*, *42*, 208-228.

Mak, M. H., Curtis, A. J., Rodd, J. M., & Gaskell, M. G. (2023). Episodic memory and sleep are involved in the maintenance of context-specific lexical information. *Journal of Experimental Psychology: General, 152,* 3087-3115.

Malmberg, K. J., Holden, J. E., & Shiffren, R. M. (2004). Modeling the effects of repetitions, similarity, and normative word frequency on old-new recognition and judgments of frequency. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*, 319-331.

McMurray, B., Horst, J. S., & Samuelson, L. K. (2012). Word learning emerges from the interaction of online referent selection and slow associative learning. *Psychological review*, *119*, 831-877.

Monsell, S., Doyle, M. C., & Haggard, P. N. (1989). Effects of frequency on visual word recognition tasks: Where are they?. *Journal of Experimental Psychology: General*, *118*, 43-71.

Morrison, C. M., & Ellis, A. W. (1995). Roles of word frequency and age of acquisition in word naming and lexical decision. *Journal of experimental psychology: learning, Memory, and cognition*, *21*, 116-133.

Murray, W. S., & Forster, K. I. (2004). Serial mechanisms in lexical access: the rank hypothesis. *Psychological Review*, *111*, 721-756.

Oldfield, R. C., & Wingfield, A. (1965). Response latencies in naming objects. *Quarterly Journal of Experimental Psychology*, *17*, 273-281.

Oppenheim, G. M., Dell, G. S., & Schwartz, M. F. (2010). The dark side of incremental learning: A model of cumulative semantic interference during lexical access in speech production. *Cognition*, *114*, 227-252.

Raviv, L., Lupyan, G., & Green, S. C. (2022). How variability shapes learning and generalization. *Trends in Cognitive Sciences*, *26*, 462-483.

Rodd, J. M., Cai, Z. G., Betts, H. N., Hanby, B., Hutchinson, C., & Adler, A. (2016). The impact of recent and long-term experience on access to word meanings: Evidence from large-scale internet-based experiments. *Journal of Memory and Language*, *87*, 16-37.

Rodd, J. M., Cutrin, B. L., Kirsch, H., Millar, A., & Davis, M. H. (2013). Long-term priming of the meanings of ambiguous words. *Journal of Memory and Language*, *68*, 180-198.

Roediger, H. L. & Butler, A. C. (2011). The critical role of retrieval practice in long-term retention. *Trends in Cognitive Sciences, 15,* 20-27.

Scarborough, D. L., Cortese, C., & Scarborough, H. S. (1977). Frequency and repetition effects in lexical memory. *Journal of Experimental Psychology: Human Perception and Performance*, *3*, 1-17.

Shiffrin, R. M., & Steyvers, M. (1997). A model for recognition memory: REM—retrieving effectively from memory. *Psychonomic Bulletin & Review*, *4*, 145-166.

Slamecka, N. J., & Graf, P. (1978). The generation effect: Delineation of a phenomenon. *Journal of Experimental Psychology: Human Learning and Memory*, *4*, 592-604.

Stark, K., van Scherpenberg, C., Obrig, H., & Abdel Rahman, R. (2023). Web-based language production experiments: Semantic interference assessment is robust for spoken and typed response modalities. *Behavior Research Methods*, *55*, 236-262.

Tamminen, J., Payne, J. D., Stickgold, R., Wamsley, E. J., & Gaskell, M. G. (2010). Sleep spindle activity is associated with the integration of new memories and existing knowledge. *Journal of Neuroscience*, *30*, 14356-14360.

Tsuboi, N., Francis, W. S., & Jameson, J. T. (2021). How word comprehension exposures facilitate later spoken production: implications for lexical processing and repetition priming. *Memory*, *29*, 39-58.

Underwood, B. J. (1969). Some correlates of item repetition in free-recall learning. *Journal of Verbal Learning and Verbal Behavior*, *8*, 83-94.

Van Heuven, W. J., Mandera, P., Keuleers, E., & Brysbaert, M. (2014). SUBTLEX-UK: A new and improved word frequency database for British English. *Quarterly Journal of Experimental Psychology*, *67*, 1176-1190.

Vitkovitch, M., & Tyrrell, L. (1995). Sources of disagreement in object naming. *The Quarterly Journal of Experimental Psychology*, *48*, 822-848.

Vogt, A., Hauber, R., Kuhlen, A. K., & Rahman, R. A. (2021). Internet-based language production research with overt articulation: Proof of concept, challenges, and practical advice. *Behavior Research Methods*, *54*, 1954-1975.

Zormpa, E., Brehm, L. E., Hoedemaker, R. S., & Meyer, A. S. (2019). The production effect and the generation effect improve memory in picture naming. *Memory*, *27*, 340-352.

**Acknowledgements**

Part of this work was funded by a Small Grant from the Experimental Psychology Society awarded to RC. We thank Annelies van Wijngaarden for collecting data for Experiment 1 and Elif Yildiz and Bilge Guney for collecting data and assisting with annotations for Experiment 2.

## Appendices

**Appendix A: Full model formula and model output for the linear mixed effects model analysis for Experiments 1 and 2.**

Table A1: Full model formula and output for data analyses in Experiment 1. RE variance = random effects variance; (p) stands for random effects by participants; (i) stands for random effects by items. All predictors are defined in the Data Analysis section.

| Session 1: Response latencies ~ Frequency * Presentation + (1 + Frequency * Presentation ‖ Participant) + (1\|Picture) | | | | |
|---|---|---|---|---|
| | *b* | *SE* | *t/p* value | RE variance |
| Intercept | 1035.85 | 25.68 | 40.33 | (p) 26089; (i) 1919 |
| Frequency | -68.11 | 12.78 | -5.33 | (p) 418 |
| Presentation | -27.93 | 2.98 | -9.36 | (p) 351 |
| Frequency * Presentation | 15.37 | 1.50 | 10.24 | (p) 4 |
| Presentation 1: Frequency | 52.73 | 12.30 | < .001 | - |
| Presentation 2: Frequency | 37.60 | 11.90 | .002 | - |
| Presentation 3: Frequency | 21.99 | 11.70 | .06 | - |
| Presentation 4: Frequency | 6.62 | 11.80 | .57 | - |
| Presentation 5: Frequency | -8.76 | 12.00 | .47 | - |
| Presentation 6: Frequency | -24.13 | 12.40 | .05 | - |
| Session 2: Response latencies ~ Frequency * Item Type + (1 + Frequency + Item Type ‖ Participant) + (1 \| Picture) | | | | |
| | *b* | *SE* | *t* | RE variance |
| Intercept | 999.54 | 28.78 | 34.74 | (p) 29304; (i) 6251 |
| Frequency | -41.28 | 15.07 | -2.74 | (p) 7 |
| Item Type | -113.78 | 15.48 | -7.35 | (p) 477 |

| | | | | |
|---|---|---|---|---|
| Frequency * Item Type | 66.64 | 30.12 | 2.21 | - |
| New items: Frequency | 74.60 | 21.80 | <.001 | - |
| Old items: Frequency | 7.96 | 20.80 | .70 | - |

Session 3: Response latencies ~ Frequency * Item Type + (1 + Frequency * Item Type || Participant) + (1 | Picture)

| | *b* | *SE* | *t* | RE variance |
|---|---|---|---|---|
| Intercept | 995.84 | 26.46 | 37.64 | (p) 25633; (i) 4739 |
| Frequency | -35.86 | 13.49 | -2.66 | (p) 436 |
| Item Type | -58.52 | 13.23 | -4.24 | (p) 169 |
| Frequency * Item Type | 51.57 | 26.49 | 1.95 | (p) 750 |
| New items: Frequency | 61.70 | 19.20 | .001 | - |
| Old items: Frequency | 10.10 | 18.60 | .59 | - |

Table A2: Full model formula and output for data analyses in Experiment 2. RE variance = random effects variance; (p) stands for random effects by participants; (i) stands for random effects by items. All predictors are defined in the Data Analysis section.

Session 1: Response latencies ~ Frequency * Presentation + (1 + Frequency + Presentation || Participant) + (1 + Presentation || Word)

| | *b* | *SE* | *t/p* value | RE variance |
|---|---|---|---|---|
| Intercept | 730.79 | 17.55 | 41.65 | (p) 11314; (i) 1767 |
| Frequency | -24.14 | 11.33 | -2.13 | - |
| Presentation | -17.75 | 1.15 | -15.41 | (i) 66 |
| Frequency * Presentation | 3.22 | 2.33 | 1.38 | (p) 4 |
| Presentation 1: Frequency | 20.86 | 11.20 | .06 | - |
| Presentation 2: Frequency | 17.57 | 11.40 | .12 | - |
| Presentation 3: Frequency | 14.29 | 11.90 | .23 | - |
| Presentation 4: Frequency | 11.01 | 12.80 | .39 | - |

| | | | | |
|---|---|---|---|---|
| Presentation 5: Frequency | 7.73 | 13.90 | .56 | - |
| Presentation 6: Frequency | 4.44 | 15.20 | .77 | - |

Session 2: Response latencies ~ Frequency * Item Type + (1 + Frequency ‖ Participant) + (1 ‖ Picture)

| | *b* | *SE* | *t/p* value | RE variance |
|---|---|---|---|---|
| Intercept | 1053.56 | 22.15 | 47.56 | (p) 15573; (i) 9258 |
| Frequency | -55.32 | 18.28 | -3.03 | (p) 389 |
| Item Type | 41.94 | 17.98 | 2.33 | - |
| Frequency * Item Type | -73.77 | 35.97 | -2.05 | - |
| New items: Frequency | 92.20 | 26.0 | <.001 | - |
| Old items: Frequency | 18.40 | 25.30 | .47 | - |

**Appendix B: Number of trials discarded for Experiments 1 and 2.**

Table B1: Number of trials discarded for high and low frequency words on each presentation in Session 1, and for old and new items in Sessions 2 and 3 in Experiment 1.

| Presentation | Frequency | Discarded trials | Analysed trials |
|---|---|---|---|
| Session 1 | | | |
| 1 | High | 177 (6.58%) | 1167 (43.42%) |
| | Low | 338 (12.57%) | 1006 (37.43%) |
| 2 | High | 104 (3.87%) | 1240 (46.13%) |
| | Low | 181 (6.73%) | 1163 (43.27%) |
| 3 | High | 105 (3.91%) | 1239 (46.09%) |
| | Low | 165 (6.14%) | 1179 (43.86%) |
| 4 | High | 114 (4.24%) | 1230 (45.76%) |
| | Low | 160 (5.95%) | 1184 (44.05%) |
| 5 | High | 109 (4.06%) | 1235 (45.95%) |
| | Low | 154 (5.73%) | 1190 (44.27%) |
| 6 | High | 121 (4.50%) | 1223 (45.50%) |
| | Low | 154 (5.73%) | 1190 (44.27%) |
| Session 2 | | | |
| Old | High | 119 (4.89%) | 1097 (45.11%) |
| | Low | 173 (7.11%) | 1043 (42.89%) |
| New | High | 197 (8.10%) | 1019 (41.90%) |
| | Low | 602 (24.75%) | 614 (25.25%) |
| Session 3 | | | |
| Old | High | 169 (6.77%) | 1079 (43.23%) |
| | Low | 159 (6.37%) | 1089 (43.63%) |
| New | High | 99 (3.97%) | 1149 (46.03%) |
| | Low | 515 (20.63%) | 733 (29.37%) |

Table B2: Number of trials discarded for high and low frequency words on each presentation in Session 1, and for old and new items in Sessions 2 in Experiment 2.

| Presentation | Frequency | Discarded trials | Analysed trials |
|---|---|---|---|
| Session 1 | | | |
| 1 | High | 17 (0.65%) | 1295 (49.35%) |
| | Low | 63 (2.40%) | 1249 (47.60%) |
| 2 | High | 9 (0.34%) | 1303 (49.66%) |
| | Low | 13 (0.50%) | 1299 (49.51%) |
| 3 | High | 7 (0.27%) | 1305 (49.73%) |
| | Low | 8 (0.31%) | 1304 (49.70%) |
| 4 | High | 7 (0.27%) | 1305 (49.73%) |
| | Low | 3 (0.11%) | 1309 (49.89%) |
| 5 | High | 7 (0.27%) | 1305 (49.73%) |
| | Low | 4 (0.15%) | 1308 (49.85%) |
| 6 | High | 8 (0.31%) | 1304 (49.70%) |
| | Low | 7 (0.27%) | 1305 (49.73%) |
| Session 2 | | | |
| Old | High | 157 (6.46%) | 1059 (43.54%) |
| | Low | 239 (9.83%) | 977 (40.17%) |
| New | High | 179 (7.36%) | 1037 (42.64%) |
| | Low | 550 (22.62%) | 666 (27.39%) |

**Appendix C: Results for the property judgement times in Experiment 2**

All 43 participants from Session 1 completed the property judgement task. Judgement times were calculated as the difference between picture onset and button press time. Before analysis, we discarded 33 trials (0.20%) shorter than 200 ms because they were too fast to reflect reasonable responses. Additionally, we discarded 51 trials (0.31%) longer than 4000 ms because they likely indicated that participants were not paying full attention to the task. We also dropped 1494 trials (9.04%) where participants responded incorrectly. Finally, we focused our analysis on the *yes* trials (9769 datapoints) because cognitive processes underlying negative (i.e., *no*) responses are more complex than those underlying positive responses.

Property judgement times to the second word were predicted by the Frequency of the first word, which participants named. Frequency was contrast coded (reference level: low vs. high; -0.5, 0.5). We also included Presentation and its interaction with Frequency in the analysis. The property for judgement was different each time, and so the Presentation number was based on the number of times the first word had been presented previously. Models fitted using the maximal random effects structure returned a singular fit error. As a result, the final model included by-item random effects for Presentation and by-participant random effects for Frequency and Presentation but not their interaction. Correlations among random effects were fit to zero to aid model convergence.

There was no difference in property judgement times for properties related to HF ($M$ = 917 ms) or LF words ($M$ = 944 ms; $b$ = -29.78, $SE$ = 24.35, $t$ = -1.22). However, judgement times decreased with each Presentation (Presentation 1 $M$ = 1042 ms; Presentation 2 $M$ = 964 ms; Presentation 3 $M$ = 923 ms; Presentation 4 $M$ = 902 ms; Presentation 5 $M$ = 869 ms; Presentation 6 $M$ = 897 ms; $b$ = -27.85, $SE$ = 5.66, $t$ = -4.92). There was no interaction between Frequency and Presentation ($b$ = 2.11, $SE$ = 4.05, $t$ = 0.52).

Figure A1. Distribution (and individual data points) of property judgement latencies for high and low frequency words on each presentation.