Interactive effects of word frequency and masked repetition in the lexical decision task: A large scale online study.

Roberto Petrosino[1], Diogo Almeida[1]

ARTICLE HISTORY
Compiled September 15, 2023

[1] Psychology, New York University Abu Dhabi, Abu Dhabi, United Arab Emirates

ABSTRACT
Masked repetition priming is believed to reflect lexical access in visual word recognition. It is reportedly word-specific and insensitive to frequency, although the latter claim is disputed. We hypothesize that past disparate findings are due to lack of statistical power. Consequently, we ran two large masked priming experiments online. The first established the feasibility of such experiments, revealing word-specific repetition effects for both high and low frequency words. The interaction between repetition and frequency was small and not statistically significant. The second study was pre-registered to have high power for detecting a 15ms interaction effect. The results again showed masked repetition priming effects, now with a 12ms interaction being statistically detected. We conclude that the repetition effect in masked repetition priming is moderately modulated by frequency. Addionally, we demonstrate the feasibility of conducting masked priming experiments online, allowing the detection of smaller effects that have historically eluded lab-based experiments.

KEYWORDS
masked priming; frequency effect; frequency attenuation; online browser-based experiment; RT; pre-registration; power analysis

1. Introduction

The masked priming technique has been an invaluable tool in visual word recognition research. It has allowed researchers to study the conditions under which orthographic, phonological, morphological, and semantic manipulations may affect access to visual word forms in a way that arguably mitigates strategic effects and minimizes the influence of controlled processes (Forster 1998). Pioneered by Evett and Humphreys (1981), this technique involves a forward mask (i.e., usually a string of hashes, #####), followed by a prime string presented for very short time ($SOA < 60$ ms), and a target string presented immediately after. Because the prime presentation is so brief and masked by preceding and subsequent stimuli, most participants report not being aware that a prime string was even presented, and can at most report a screen flicker just before the target presentation (Forster, Mohan, and Hector 2003). This technique has been

CONTACT: Roberto Petrosino. Email: roberto.petrosino@nyu.edu. Diogo Almeida. Email: diogo@nyu.edu.

extensively used in the last fifty years as a tool to answer questions about the mental lexicon, and word recognition processes. In general, three main models of priming can be identified. In the persisting activation model (McClelland and Rumelhart 1981; Grainger and Jacobs 1996; Coltheart et al. 2001), priming is seen as the "head start" effect that the prior presentation of the prime gives to the activation of the target. In memory recruitment models (a.o., Bodner and Masson 1997; Masson and Bodner 2003), priming is seen as the effect whereby the prior recruitment of the memory representation of the prime assist with the identification of the memory representation of the target. Finally, in the entry-opening models (Forster and Davis 1984; Forster 1998, 1999), the priming response is seen as a "post-access effect" triggered by prior opening of the prime entry. Each of these models makes specific predictions about the potential interaction of priming effects with other relevant variables, such as prime masking (i.e., SOA), and word frequency, depending on how these two variables may be encoded within each model. In the persisting activation model, the priming magnitude is expected to be inversely proportional to word frequency, if it is encoded in terms of changes in connection strengths (or activation thresholds); but insensitive to it, if it is rather encoded in terms of variations in the resting activation level of word units. Similarly, in the memory recruitment model, the priming magnitude is expected to inversely interact with word frequency, since word frequency is encoded episodic distinctiveness. As the episodic representations of low-frequency words is more distinctive (i.e., less usually heard and used) than the representations of high-frequency words, the priming response to the former should be greater than the latter. Finally, in the entry-opening model, priming and frequency should not interact, as priming is argued to arise only after the entry of the target has been located (Forster, Mohan, and Hector 2003).

The research of the last 50 years seem to support the latter model, in reporting results that are qualitatively different in the two different masking environment. One the one hand, word repetition priming effects in the non-masked environment reportedly interact with word frequency, in that they are numerically bigger for low-frequency words than for high-frequency words (*frequency attenuation effects*: Scarborough, Cortese, and Scarborough 1977). On the other hand, a wealth of evidence have shown that word repetition priming effects in the masked environment are the same for low- and high-frequency words (Forster and Davis 1984, 1991; Forster et al. 1987; Segui and Grainger 1990; Sereno 1991; Rajaram and Neely 1992; Bodner and Masson 1997; Forster, Mohan, and Hector 2003), thus giving support to the entry-opening model, and the view that the frequency attenuation effects arising in the non-masked environment are due to episodic strategies, which are drastically reduced (if not completely disposed of) when the prime is sufficiently masked (i.e., when it is not presented for longer than 60 ms). As Table 1 shows, however, a closer review of the relevant literature reveal that few, more recent studies have actually reported frequency attenuation effects in the masked environment (Bodner and Masson 2001; Kinoshita 2006; Norris and Kinoshita 2008; Nievas 2010). In particular, Bodner and Masson (2001) report as statistically significant interaction between repetition masked priming and lexical frequency in a series of experiments in which the frequency gap between low and high frequency words was increased, and the target word was presented in mixed case to increase the lexical decision difficulty and favor the strategic processing of the information gleaned from the prime. Kinoshita (2006) noticed that in earlier studies the low frequency words often had very high error rates, and suggested that perhaps many participants did not know them.

Table 1.: Summary of the repetition priming effects as a function of word frequency reported in the literature. Significant interaction effects are signaled with an asterisk (*) in the Interaction column.

				MOP		
Study	Language	N	SOA	HF	LF	Interaction
Forster & Davis (1984), exp. 1	English	28	60	45	38	-7
Forster, Davis, Schoknecht, & Carter (1987), exp. 1	English	16	60	61	66	5
Forster, Davis, Schoknecht, & Carter (1987), exp. 4	English	27	60	34	25	-9
Sereno (1991), exp. 1	English	20	60	40	64	24
Forster & Davis (1991), exp. 5	English	24	60	54	72	18
Rajaram & Neely (1992), exp. 1	English	48	50	30	37	7
Rajaram & Neely (1992), exp. 2	English	48	50	45	78	33
Bodner & Masson (1997), exp. 1	English	24	60	29	45	16
Bodner & Masson (1997), exp. 3	English	24	60	36	50	14
Bodner & Masson (2001), exps. 2A, 2B, 3, & 6 (average)[1]	English	40	60	37	69	32
Forster, Mohan, & Hector (2003), exp. 1	English	24	60	63	60	-3
Kinoshita (2006), exp. 1	English	24	53	32	38	6
Kinoshita (2006), exp. 2	English	24	53	29	59	30
Norris & Kinoshita (2008), exp. 1	English	24	53	35	66	31
Segui & Grainger (1990), exp. 4	French	36	60	42	45	3
Nievas (2010), exp. 1b	Spanish	30	50	44	65	21
Nievas (2010), exp. 2a	Spanish	30	50 or 33[2]	51	58	7
Average				42	55	13

[1]Reported in Masson & Bodner (2003)
[2]SOA for each subject determined by pre-test

If participants treated a substantial number of low frequency words as non-words, it could artificially depress the repetition priming effect in that condition. In two separate experiments, Kinoshita (2006) showed that larger repetition priming effects for low frequency words were only obtained when the low frequency words were especially vetted to make sure the participants knew them prior to the experiment. Following this finding, Norris and Kinoshita (2008) were also able to find an interaction between lexical frequency and repetition in masked repetition priming. Similar results are reported in the series of masked priming experiments in Spanish reported by Nievas (2010). The study reported here aims at reconciling the conflicting evidence reported in the literature and just briefly summarized above. To this end, we identified two potential sources of such seemingly contradictory results. First, we take issue with the previous literature, in that it unanimously made use of old database (Kučera and Francis 1967), now commonly deemed outdate dand unreliable. In our study instead, all materials were based on two renowned word frequency databases: HAL (Lund and Burgess 1996) and SUBTLEX$_{US}$ (Brysbaert and New 2009).

Second, we shared the generally concern, initially voiced by both Bodner and Masson (1997) and Kinoshita (2006), that previous studies might have been underpowered to detect an interaction effect, which usually requires a much larger sample size than main effects (e.g., Potvin and Schutz 2000; Brysbaert and Stevens 2018). To this end, we conducted a power analysis and designed a properly powered experiment to detect even a small interaction effect. Of course, we were aware that ensuring an experiment to be properly powered might have meant an exponential increase of the sample size. Large-scale experiments, however, may not always be practically doable in standard, controlled environmental setting (e.g., a quite, sound-shielded room in a university lab). To overcome this practical difficulty, we ran this study entirely online, encouraged by the growing body of behavioral research being conducted online, especially in the post-pandemic era. In recent years, the number of cognitive science labs and departments capitalizing on the HTML5 capabilities has rapidly increased, also thanks to the proliferation of stable and powerful software packages for stimulus delivery and data collection, such as *jsPsych* (Leeuw 2014), *PsychoJS* (the javascript counterpart of PsychoPy, Peirce et al. 2019), and, more recently, *Labvanced* (Finger et al. 2017). Online experimentation has three main advantages: (i) paramount increase in the potentially recruitable sample size; (ii) dramatic decrease in the time needed to reach the desired sample size; and (iii) easy access to populations that may not be easily available in the immediate geographical proximities.

In what follows, we report two separate experiments, tackling two different, though related goals. Experiment 1 aims to evaluate the quality of masked priming data collected online, describe additional pre-processing steps, and ultimately assess its drawbacks. While Recent contributions have already reported the substantial comparability of online and in-lab masked priming studies (among others, Angele et al. 2022), we report a fully-fledged analysis of the data collected of a pilot experiment (Experiment 1) to empirically assess the reliability of a newly-available online platform that has never been tested with masked priming, *Labvanced* (Finger et al. 2017). This choice was motivated by the fact that it is a GUI-based web app that allows researchers to dispense with local installation issues (thus preventing potential incompatibilities with the CPU of the local machine and ensuring cross-platform consistency) and yet another programming language to learn (thus facilitating experimental design and deployment). This was also done to provide researchers in the same field a thorough report of raw data collected online on a large scale, in the hope that it may be of help for a sensible assessment of the risks and benefits that online experimentation may entail. Results from Experiment 1 show that masked identity priming may be reliably elicited online, modulo some extra care in the experiment planning and some additional outlier removal procedures to be performed prior to analysis. Experiment 2 was designed to tackle the issue of frequency attenuation effects in the masked repetition priming response by capitalizing on modern technologies to recruit online a sample size large enough to potentially detect even small effect sizes which would have little probability of detection otherwise. To this end, prior to running experiment 2 we a full-fledged power analysis to estimate the right sample size for this purpose prior to starting data collection. Our results showed a significant frequency attenuation effect, with the identity priming response to low-frequency words being twice as large as the identity priming response to high-frequency words. These results suggest that priming elicitation is indeed influence by word frequency, thus challenging the well-accepted claim that the priming response operate prior to lexical access.

2. Experiment 1

2.1. *Methods*

2.1.1. Materials

Three hundred five-letter English words were selected from the English Lexicon Project [Balota et al. (2007); henceforth, ELP], in which each third (i.e., 100 words) was selected from the upper, mid, and lower frequency range. We also made sure that the three frequency ranges were well separated from one another. The three word subsets corresponded to the three word conditions being tested: the high-frequency, mid-frequency, and low-frequency conditions. In each condition, fifty words were randomly chosen to be presented as targets, and the remaining fifty could be presented as unrelated primes. Table 2 reports the relevant descriptive statistics for each word condition.

Table 2.: Experiment 1. Descriptive statistics of the word item used.

		log HAL				length	
condition	N	min	max	mean	sd	mean	sd
high	100	10.01	11.97	10.87	0.60	5	0
mid	100	6.01	7.99	6.96	0.55	5	0
low	100	3.04	5.01	4.39	0.50	5	0

Three-hundred five-letter non-words were randomly selected from the ELP database as well. Half of them were randomly selected to be presented as targets; the other half was instead used as unrelated non-word primes.

2.1.2. Procedure

Three hundred participants were recruited on Prolific. Several criteria were selected so to ensure recruitment of native speakers of English. All participants had to be born in the Unites States of America, speaking English as their first and only language, and have no language-related disorder. We encouraged subjects to avoid any sort of distraction throughout the experiment, and to close any program that may be running in the background during the experiment, as a way to boost a decent performance of the stimulus presentation tool as much as possible. Besides, subjects could not be monitored in any way during data collection. Finally, to reduce the variability across devices, we restricted the experiment to be run on the Chrome browser only, since other browsers appears to be perform worse in previous pilot runs (possibly as a byproduct of the different engines used).

Each recruited participant was assigned one between two different word lists, which differed only in the relatedness of the prime with respect to the target; other than that, the two lists presented the same set of target words and non-words (300 in total). In one list, the four conditions (high-frequency, mid-frequency, low-frequency word conditions, and the non-word condition) had 25 target items being preceded by themselves (the *related* sub-condition) and the remaining 25 target items being preceded by one of the unrelated primes belonging to the same frequency bin (the *unrelated* sub-condition). In the other list, the order was reversed.

After being recruited, participants were asked to click on a link which redirected them to Labvanced. During the experiment, they were asked to perform a lexical decision task by pressing either the 'J' (for word) or 'F' (for non-word) keys on their keyboard. Each trial consisted of three different stimuli appear at the center of the screen: a series of hashes (#####) presented 500 ms, followed by a prime word presented for 33 ms, and finally the target word; the target word disappeared from the screen as soon as a decision was made. The motivation behind the choice of such a short prime duration (as compared to the literature, in which it is usually between 50 and 60 ms; see Table 1) is two-fold. First, several previous pilot experiments being run on the same platform showed that having a longer prime duration increased the number of trials with a prime duration above the subliminal threshold (usually set at 60 ms), which could trigger experiment-wide strategic influences onto the masked priming response. Second, setting such a short prime duration may ultimately maximize subliminal effects, thus ensuring the reliability of the results.

Subjects were also given 5 breaks throughout the experiment. When the experiment was over, the participants were then redirected to Prolific in order to validate their submission. The median time to finish the experiment was 11 minutes and 27 seconds. Each participant was paid with the standard rate of GBP 9/hour.

2.2. *Data analysis*

Analysis scripts and an abridged version of the data collected can be found on OSF (https://osf.io/rnyhp). We performed three different steps of analyses (in sequential order), with the goal of gaining a thorough understanding of the data collected (92,690 observations in total). The first step of analyses is the novel analysis step that is usually not included in the typical analysis pipeline for RT-based data, and looks at at the distribution of the actual duration of prime words for each trial for each subject. This additional analysis step is indeed necessary to fully understand the performance capabilities of the engine Labvanced relies on. The second and third step of analyses are instead part of the typical analysis pipeline for RT-based data, looking at subject performance and RT distribution, respectively.

2.2.1. *Prime time*

During the experiment, the duration of presentation of the prime word was recorded online for every trial, as an additional measure necessary for a thorough assessment of the stimulus delivery engine in terms of reliability and variance of the duration of the presentation of the stimuli, and in particular of the prime. The distribution of the prime durations recorded in the experiment are shown in Figure 1 below. Both the mean (mean = 37.71) and the median (median = 35) of the prime duration was slightly greater than the target prime duration (33 ms). Overall, the distribution showed a positive skew and a long right tail.

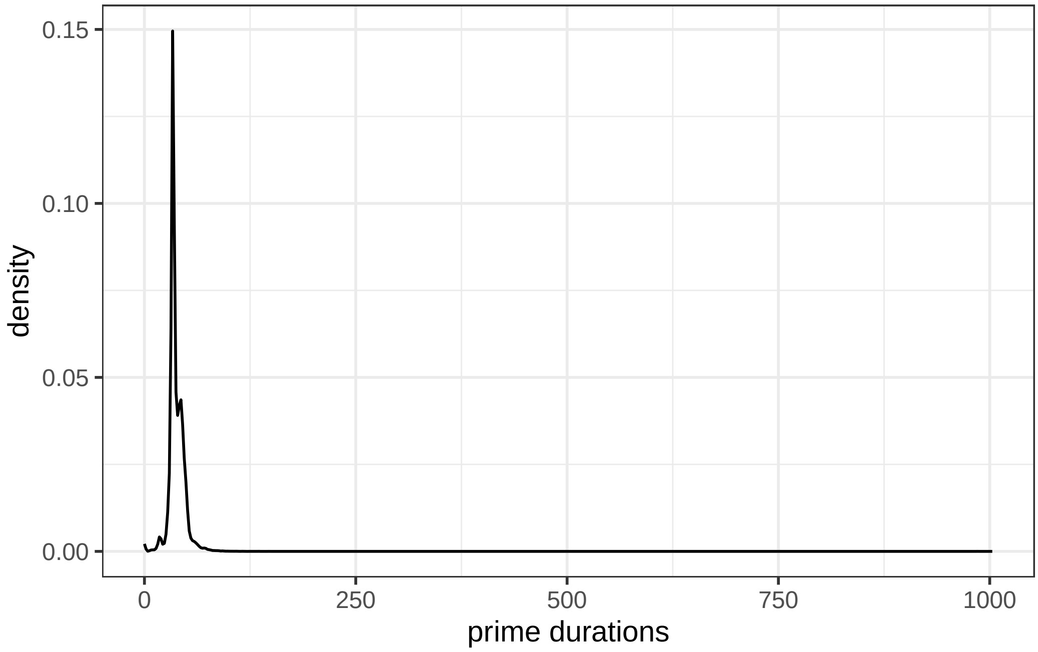

Figure 1.: Experiment 1. Distribution of the durations of the primes across trials and subjects. The red and green vertical lines represent the mean and median prime durations, respectively.

The distribution above suggests that, while overall the web engine presented most trials at the preset duration, it was not as precise and accurate as a local engine. This was expected and likely due to the great variation in the specifications of the devices used by the participants, and may likely be impossible to control, at least at the current state of development of the online platforms available at this time. However, in masked priming, in which the duration of the prime is essential part of the design itself, such fluctuations may indeed hinder proper elicitation of the response. As a way to counteract the potential influence that such fluctuations might have had on the priming response, we only kept trials whose prime durations were within a pre-set range from the target duration of 33 ms. Taking a standard 60-Hz monitor as reference, the lower and the upper bounds was set at half of a full refresh cycle (i.e., 8 ms) and at 60 ms respectively, so to keep trials that would not fall below or above the subliminal threshold. Out of the 92,690 observations collected, only 3.08% of the trials were out of the range selected, the great majority of which (3.57%) were above the range set. We take this as further corroborating the argument that Labvanced is capable to reliably present stimuli at short durations. However, by way of ensuring the quality of the data collected and therefore reliability of the results, recording of the actual durations of the prime stimuli of each trial and removal of the out-of-range trials as a cautionary, preliminary step in the analysis pipeline is recommended.

Table 3.: Experiment 1. Breakdown of the trials in which the prime word was presented in, above, or below the preset time range (25<33<49 ms).

range	trials	% trials
above	102	0.11
below	2754	2.97
in range	89834	96.92

Labvanced provides an additional measure, the callback loop measure, which assesses the precision and accuracy of the machine the experiment is being run one. For each subject, and completely separate from the logic of the experiment being run, the time required for the loop being presented is stored as a timestamp every 5,000 ms (set as a time out call) and the difference between the two values (the time out, i.e. 5,000 ms, and the actual timestamp). The mean and the standard deviation of all these differences are then calculated. Nearly 35% participants had a mean above 10 ms and a standard deviation above 5 ms, and were removed from analysis. The distribution of the prime durations was then re-assessed, and revealed that the callback loop measure was marginally useful, as it reduced the number of out-of-range trials to 2.55%. However, in the effort to ensure high quality data, both cutoffs (the preset time duration cutoff and the callback loop cutoff). These preliminary steps meant that about 34.78% of the initial sample size had to be removed due to performance and stimulus timing delivery issues. A total of 195 participants and 57,138 observations were included in the next steps of analysis.

2.2.2. Subject and item performance

Non-word trials were excluded from analysis a priori, as not strictly relevant to the specific question being asked in the experiment. The by-item word error rate revealed that 36 words had an error rate higher than 30%, and they all belonged to the low-frequency condition. The high number of words with a high error rate was likely due to their low frequency (see Section 2.1.1 for further information), so we decided to remove the low-frequency condition altogether, since removing the whole set of high-error words (roughly corresponding to 72% of the total items of the low-frequency conditions) would have drastically decreased of the number of observations for the low-frequency condition, thus impinging on the reliability of the estimates of that condition. After removing the entire low-frequency condition, we re-calculated word error rates. only 3 mid-frequency words (*carte, parse, posit*) had an error rate higher than 30%, and were removed from analysis. We then calculated the subject error rates. Only 1 subject was removed because of their overall error score was higher than 30%. After this cut, we also calculated the sensitivity measure d (Green and Swets 1966) from the by-subject error scores to assess participants' attentiveness to the lexical decision task. The measure d' is usually calculated as the difference between the by-subject z-transformed percentages of hit (i.e., a word correctly recognized as a word) and false alarm (i.e., a non-word incorrectly recognized as a word) scores. A d' value close to zero generally indicates a lack of attentiveness/awareness of the participant onto the stimulus.

The distribution of d' across participants was never below 1.5, which suggested that all participants were actively engaging with the task. Finally, we made sure that all subjects included in the analysis had at least half of the stimuli that were presented with for each condition*primetype subdesign (i.e., 25); therefore, subjects with less than 12 stimuli in at least for one condition were discarded. This was an additional cautionary step to avoid inaccurate estimates due to a low number of trials. Only 1 subjects were removed for this reason. After removing the incorrect responses, a total of 17,310 observations and 193 participants were included in further analyses.

2.2.3. RT distribution

Finally, individual trials were excluded if the relative RT was below 200 ms and 1800 ms. 71 observations were excluded at this stage of analysis (0.41% of the dataset), which led to a total of 17,239 observations that were actually included in the statistical analysis below.

2.3. ***Results***

For each condition, priming effects were calculated for each subject by subtracting the by-subject mean RT to the related sub-condition from the by-subject mean RT to the unrelated sub-condition. Standardized effect sizes (i.e., Cohen's *d*) were then calculated for each condition. Finally, by-subject mean priming effects were grand-averaged across subjects for each condition. Figure 2 and Table 4 below report the descriptive statistics of the experiment. Both word conditions show non-null priming effects. The priming magnitudes (and the relative standardize effect sizes) are different in terms of magnitude and sign. The high and mid-frequency word conditions show similar positive effects, with the mid-frequency priming effects being 6-ms greater that the high-frequency priming effects.

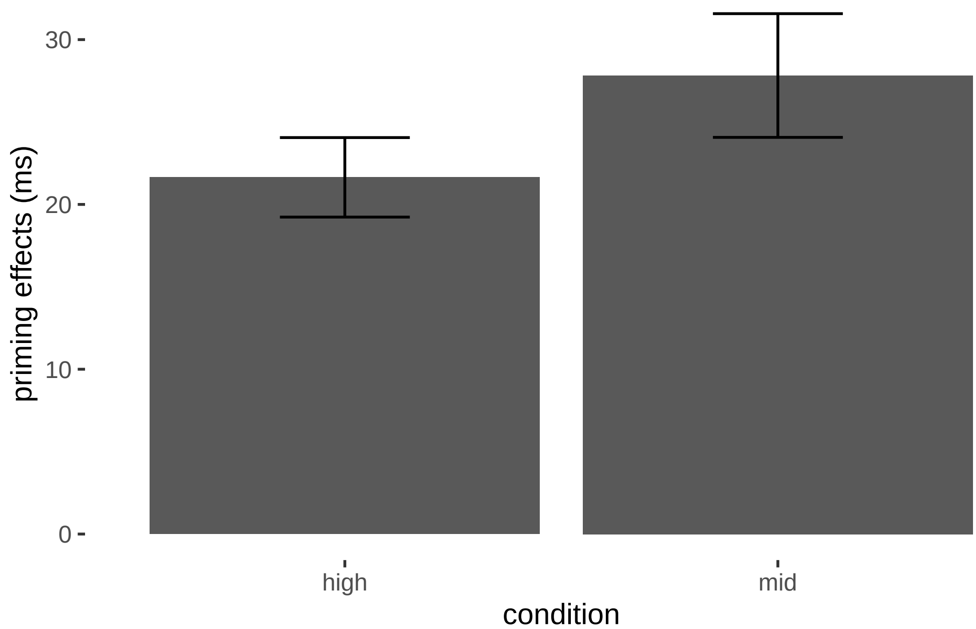

Figure 2.: Experiment 1. Barplot of the priming effects (error bars represent one standard error in either direction)

Table 4.: Experiment 1. Mean RTs of the related and unrelated conditions and Pearson's r between them, priming magnitudes, standard deviation (SD), and Cohen's d (ES).

condition	related	unrelated	cor	priming	SD	ES
high	596	618	0.92	22	34	0.65
mid	666	694	0.85	28	52	0.53

A two-way ANOVA was performed, in which RT was the dependent variable, and *primetype* (2 levels: unrelated vs. unrelated) and *condition* (2 levels: high vs. low) were the two independent variables. Our analysis showed significant main effects for both independent variables but a trending-to-significant interaction effects. The specifics of the model are reported in Table 5.

Table 5.: Experiment 1. Summary of the statistical results

term	t	SE	p-value
intercept (high/unrelated)	263.62	2.34	< 2e-16
condition (mid)	22.76	3.45	< 2e-16
primetype (related)	−6.52	3.31	< 2e-16
condition * primetype	−1.71	4.85	.087

2.4. ***Discussion***

The main purpose of experiment 1 was to present results of a typical masked repetition priming experiment carried out in large scale ($N > 100$) online, and assess whether modern online stimulus delivery programs such as *Labvanced* are able to provide data that are comparable in quality to data collected in-lab. Our extensive analysis of the data of experiment 1 suggested that *Labvanced*, as much as any similar online stimulus delivery programs, is indeed able to provide data of acceptable quality from a larger, potentially infinite sample size, even though a few additional cautionary steps are recommended to ensure reliable results. In particular, we identified one main issue that seems unavoidable, at least at the current stage of program development and machine capacities: a fundamental imprecision and inaccuracy in the stimulus duration. While time precision and accuracy might be negligible in other experimental designs, it is indeed a crucial concern in masked priming design, where the correct interpretation of the masked priming response is contigent on the accurate and precise duration of the presentation of the prime. The number of potential factors impinging on this are diverse, and factually not controllable: e.g., variability of the devices being used by participants during the experiment, internet connection speed, and number of programs running in the background, just to name a few. Such time fluctuations in the prime presentation are a primary concern, which led us to remove trials with a prime duration that was beyond the subliminal thresholds (28-60 ms). This meant that overall around 35% had to be discarded from analysis.

Regardless of such methodological issues, the data was still able to elicit significant masked priming effects for both the high- and mid-frequency conditions. However, similarly to what most literature reported in the past, the frequency attenuation effect (i.e., the difference effect size between the low- and high-frequency priming responses) was numerically not-null (around 6 ms), but only trending to significance. As suggested elsewhere (Potvin and Schutz 2000), the lack of a significant interaction effect might have just been due to low statistical power. We address this concern by running a full-fledged power analysis to quantify a feasibly sample size that would guarantee acceptable statistical power (>80%) to catch the potential interaction between priming and frequency in experiment 2.

3. Experiment 2

From a purely descriptive perspective, the status of frequency attenuation effects in masked priming in quite unequivocale. Out of the seventeen studies listed in Table 1, fourteen have numerically larger repetition effect sizes for low frequency words compared to high frequency words; and only the remaining three report results in the opposite direction. Nevertheless, only three experiments report a significant frequency attenuation effect, whose size is in the 30 ms range, which is about as large as the size of the masked priming repetition priming effects to high-frequency. Coincidentally, around half of such range (i.e., 13 ms) is the average frequency attenuation effect across all studies. These considerations seems to suggest that there may be a real frequency attenuation effect in masked priming, but it is smaller than what most experiments conducted so far might have been powered to detect. Building on the encouraging results of experiment 1, experiment 2 directly addresses the statistical power issue by leveraging the capabilities of online data collection. To this end, before running the actual experiment, we performed an extensive power analysis to estimate a sample size large enough to detect the expected interaction effects (see below and the Supplemental materials for further details).

3.1. ***Methods***

3.1.1. Preregistration

As an additional effort to transparency, we preregistered the results of the power analysis, the goals, and the design and analysis plan for this experiment prior to data collection. This study's desired sample size, included variables, hypotheses, and planned analyses were preregistered on Open Science Framework (https://doi.org/10.17605/OSF.IO/3NFQP) prior to any data being collected.

3.1.2. Materials

Materials for experiment 2 were prepared in a slightly different way than the materials for experiment 1. First, only two frequency conditions were tested: a low-frequency condition and a high-frequency condition, while ensuring they included all words likely being known by participants. Second, words were still sampled from the English Lexicon Project [ELP; Balota et al. (2007)], but based on the SUBTLEX$_{US}$ frequency index, rather than the HAL frequency index, as suggested by Brysbaert and New (2009). For each condition, two hundred four-letter English words were sampled. We were forced to reduce the orthographic length of the item by one letter to maintain a clear-cut distinction between the frequency ranges of the two word conditions. One hundred words were sampled from the upper bound frequency range (between 3.9 and 5.5 in the log SUBTLEX$_{US}$ frequency scale). One hundred words were sampled from the lower bound frequency range (between 1.5 and 2 in the log SUBTLEX$_{US}$ frequency scale). We also ensured that the low-frequency words were known to the average English speaker. From each word set, fifty words were randomly chosen to be presented as targets, so that the remaining fifty could be presented as unrelated primes. We made sure that all words used were monomorphemic nouns, adjectives, or verbs, thus excluding particles, prepositions, and participial forms. Table 6 reports the relevant descriptive statistics for each word condition.

Table 6.: Experiment 2. Descriptive statistics of the word items used.

		SUBTLEX-US					
frequency bin	N	min	max	mean	SD	length mean	length SD
high	100	164.35	5721.18	698.18	845.24	4	0
low	100	0.55	1.92	1.26	0.41	4	0

Two-hundred four-letter non-words were randomly selected from the ELP database as well. Half of them were randomly selected to be presented as targets; the other half was instead used as unrelated non-word primes. We made sure that all non-words used did not contain any existing English morpheme.

3.1.3. Sample size rationale

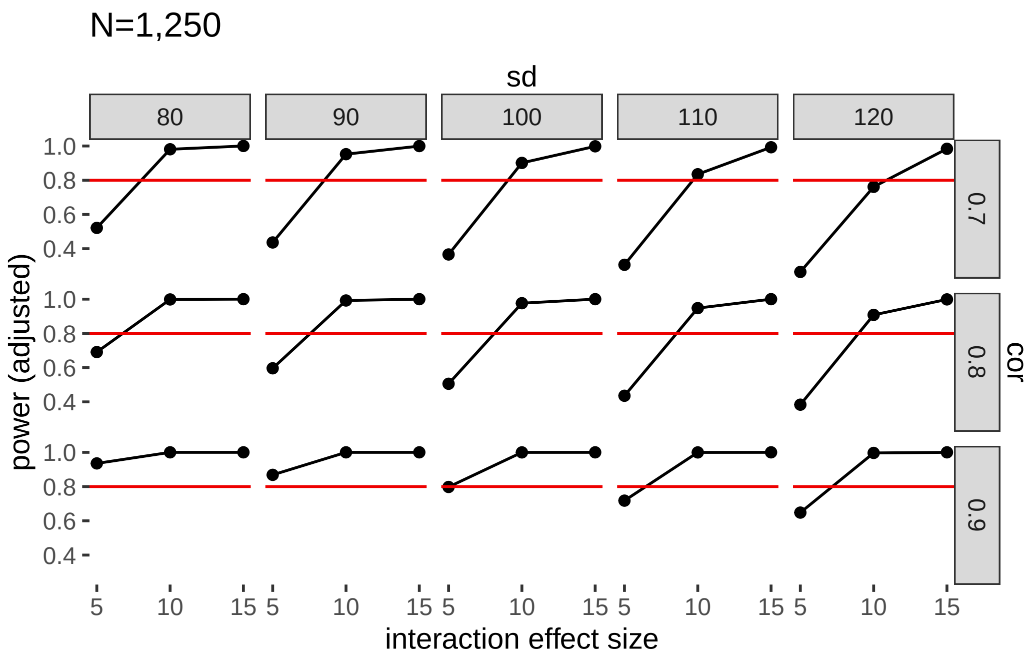

Figure 3.: Power simulations with a sample size of 1,250, for all combinations of standard deviation (sd), pairwise correlation (cor), and interaction effect size. The red line identifies the threshold of 80% power.

We ran an extensive power analysis based on the combination of four main parameters: sample size, standard deviation, pairwise correlations between the related and unrelated conditions, and interaction effect size (see **Supplemental material** for an extensive description and the the script used). These parameters were used to simulate datasets.

For each dataset, we performed a dependent-samples t-test, and then calculated power as the average of the number of significant tests (i.e., with $p < \alpha$) obtained. Out of the parameters aforementioned, only the sample size and the raw effect sizes are consistently reported in published studies, whereas the other two are very seldom reported, if ever. For this reason, our power analyses required testing multiple combinations of the four parameters. We opted for range between 60 and 100 ms (with 10 ms increments) for the standard deviation, and between 0.7 and 0.9 (with 0.1 increments) for the correlation. As for the sample size, the literature on large-scale online studies is relatively newborn, so we simulated data for a wide range of sample size between 200 and 3,000 participants (with 150 unit increments). Finally, some adjustments needed to be done for the expected (raw) effect size. All previous studies used an SOA ranging between 50 and 60 ms, whereas our previous attempts to collect masked priming data online showed that, in order to get the majority of primes to fall within the subliminal processing window (i.e., 28-60 ms), a SOA of 33 ms is recommended (rather than 49 ms, assuming a refresh rate of 16 ms, as in standard 60-Hz monitors; see Section 2.1.2). This difference must be taken into account, since the masked repetition priming effect size seems to be a function of SOA (Forster, Mohan, and Hector 2003). In order to adapt our estimate of a true effect size for the frequency attenuation effect with a prime duration that is about half of the one used in the aforementioned studies (33 ms), we performed power simulations assuming an interaction effect size of 15 ms, 10 ms and 5 ms. The first effect size (15 ms) is about half of the ones observed in studies that had a significant interaction (~30 ms). The second effect size (10 ms) is close to the average frequency attenuation effect found in the literature review (13 ms). The last effect size (3 ms) is an estimate of a lower boundary of a theoretically interesting effect size. We identified a sample size of 1250 participants as a feasible sample size that would allow us to reach an acceptable statistical power (> 80) in most combinations of parameters, and especially with a raw interaction effect size equal to or higher than 10 ms (Figure 3), while remaining within our funding capacities. Given the known limitations in time accuracy and precision of the current online stimulus delivery programs currently available (see Section 2.2.1), the target sample size was maxed out to 2,600, so to ensure the largest sample size possible.

3.1.4. Procedure

Experiment 2 was conducted in the same way as experiment 1 (see Section 2.1.2). The median time to finish the experiment was around 9 minutes.

3.2. *Data analysis*

Analysis scripts and an abridged version of the data collected can be found on OSF (https://osf.io/4uknh), and consisted of 92,690 observations in total. We performed the same three steps of analyses described for experiment 1.

3.2.1. Prime time

Prime fluctuations were dealt with in the same way as in experiment 1, so we will gloss over the extensive analysis we provided for experiment 1. We applied the callback loop cutoffs (mean: 10 ms; sd: 5 ms), which meant that around 30% of the participants had to be removed from analysis, We then applied the prime duration cutoff (25-60ms) set for experiment 1, which removed 8% of the trials. A total of 1766 participants and 334,582 observations were included in the next steps of analysis.

3.2.2. *Subject and item performance*

We calculated by-subject overall error rates (that is, including words and non-words) and remove subjects whose overall error scores were higher than 30%. 881 subject was removed because of this. It is unclear to use the cause for such a high number of low performers as compared to experiment 1 [and more generally to our previous online studies; see Petrosino (2020), Petrosino, Sprouse, and Almeida (submitted)]. At a closer look, we found that the performance of some subjects was well below chance. While the strict selection criteria set on Prolific usually guarantee high-performing participants, the large sample size required might have led to recruit less cooperative participants. After this cut, *d'* was never below 1.5. As for the word error rate, 8 mid-frequency words (*gala, glib, mitt, pang, prod, spec, tram, veer*) had an error rate higher than 30%, and were removed. Similarly to experiment 1, non-words were not analyzed, as being not relevant to the specific question being asked here. Finally, we made sure that all subjects included in the analysis had at least half of the stimuli that were presented with for each condition*primetype subdesign (i.e., 25). 12 subjects were removed for this reason. After removing the incorrect responses, a total of 71,394 observations and 873 participants were included in further analyses.

3.2.3. *RT distribution & power evaluation*

Finally, individual trials were excluded if the relative RT was below 200 ms and 1800 ms, similarly to what we did for experiment 1. 485 observations were excluded at this stage of analysis (0.68% of the dataset), which led to a total of 70,909 observations and 873 subjects that were actually included in the statistical analysis below.

Table 7.: Experiment 2. Standard deviation (SD) and pairwise correlation (cor) for each condition of the dataset.

condition	cor	SD
high	0.91	87
low	0.87	110

Such sample size is about 30% smaller than our target sample size, which could mean a substantial decrease in power. Before recruiting more participants, we evaluated the power of a simulated dataset with the same sample size (873), and standard deviation and pairwise correlation as ours (Table 7) by spline interpolation with the `zoo` R-package (Zeileis and Grothendieck 2005). Figure 4 shows that, should the raw interaction effect size be equal or higher than 10 ms, the sample size recruited would be able to ensure an acceptable statistical power (i.e., > 80), egardless of the potential variations in standard deviation and correlation across conditions. Given that most of previous studies have reported a raw interaction effect size higher than 10 ms, we therefore decided not to collect further data.

3.3. ***Results***

For each condition, priming effects were calculated in the same way as experiment 1. Figure 5 and Table 8 below report the descriptive statistics of the experiment. Both word conditions (high and low-frequency word conditions) showed non-null priming effects,

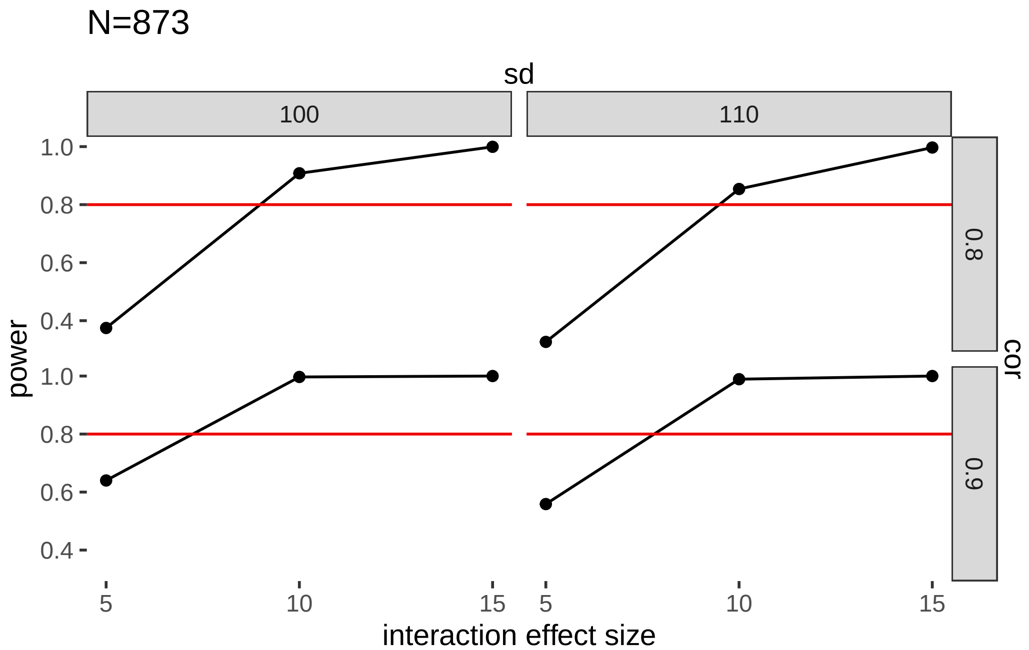

Figure 4.: Experiment 2. Power estimates for the actual sample size recruited after the outlier removal pipeline, given the actual standard deviation (sd) and pairwise correlation (cor) of the dataset.

with the low-frequency priming effects being bigger than the high-frequency priming effects.

Table 8.: Experiment 2. Mean RTs of the related and unrelated conditions, priming magnitudes, standard deviation (SD), Pearson's r between the related and unrelated conditions, and Cohen's d (ES).

condition	related	unrelated	cor	priming	SD	ES
high	567	583	0.91	16	38	0.42
low	673	703	0.87	30	56	0.54

The statistical analysis was the same as the one of experiment 1. The two-way ANOVA showed significant main effects for both independent variables, and a significant interaction effects. The specifics of the model used are reported in Table 9.

Table 9.: Experiment 2. Summary of the statistical results

term	t	SE	p-value
intercept (high/unrelated)	507.28	1.15	$<$ 2e-16
condition (low)	69.80	1.75	$<$ 2e-16
primetype (related)	−10.21	1.62	$<$ 2e-16
condition (low) * primetype (related)	−6.01	2.49	1.86-09

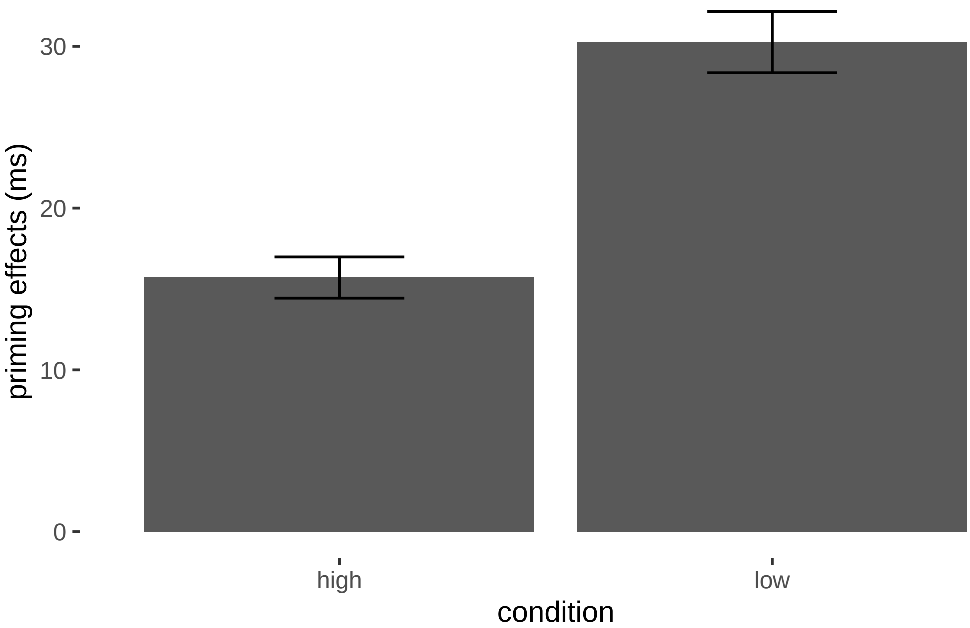

Figure 5.: Experiment 2. Barplot of the priming effects (error bars represent one standard error in either direction)

3.4. *Discussion*

Experiment 2 was specifically designed to to test the extent to which frequency attenuation effect–an pretty stable effect in the non-masked environment (i.e. with a SOA $> 60ms$)–could also be elicited in the masked environment (with SOA $< 60ms$), with a sample size large enough to reach an acceptable statistical power. To this end, we first ran a power analysis that would allows us to quantify the right number of participants to recruit to ensure good power. We identified 1,250 participants as an acceptable size, but we maxed out the size to 2,600 subjects so to accommodate for the known timing issues that come with online data collection. We therefore presented native English speakers with a lexical decision task, in which words from two distinct frequency ranges in the SUBTLEX$_{US}$ were presented in a typical masked priming fashion. Even though around 35% of the participants were removed from the statistical analysis, we estimated that the final sample size (873) could still meet our power needs, should we obtained a raw interaction effect size equal to or higher than 10 ms. As expected, our results confirmed significant main effects for frequency and priming. Moreoever, we also found a significant interaction effect, with the low-frequency condition triggering priming effects roughly twice as large as the high-frequency condition. These results clearly suggest that frequency *does* modulate the priming response, in contrast with what has been believed so far. The main theoretical implication of this experiment is that the mechanics of masked priming may indeed involve access to the mental lexicon – at least, to some degree. An in-depth discussion of such an important result is deferred in the general discussion.

4. General discussion

This paper aims to resolve the apparent contradiction in the results reported in the masked priming literature on frequency attenuation effects. While most studies report no significant interaction between frequency and the priming response, a small subset (3/17) does. In this paper, we claimed that there may be at least two sources of such contradictory results. First, all of these studies used the outdated Kučera and Francis (1967)'s word frequency database. As a consequence, any frequency-based binning procedure based on that database may be just unreliable (Brysbaert and New 2009). Second, the relevant studies suffer from a general lack of statistical power due to an insufficient sample size, especially if we consider that statistical power naturally decreases when interaction effects are considered (Brysbaert and Stevens 2018). The main consequence of low statistical power is the inability of the data to detect small, yet significant effect sizes, and therefore a substantial increase in the Type-II error rate. In the effort to shed light on the aforementioned contradicting results, the study aimed at tackling both issues by running two different experiments. To address the frequency database issue, we prepared our materials by relying on the modern database of HAL and SUBTLEX$_{US}$ (as suggested in Brysbaert and New 2009). The power issue was more complex to address. In practice, the only way to increase statistical power is to increase the sample size enough to reach an acceptable statistical power (i.e., 80%, as suggested by Cohen 1992). However, detecting small effects may require a sample size likely in the thousands of participants, which may be impractical and time-consuming in a typical in-person data collection setting. One flexible and fast solution is to take advantage of modern online stimulus delivery programs, which directly run on browsers and seem to provide comparably high data quality. Experiment 1 aimed to assess the empirical reliability of *Labvanced*, a new stimulus delivery online application that provides researchers with an intuitive GUI for the actual experimental design and data analysis, and does not presupposes any prior knowledge of a specific programming language (such as javascript). In experiment 1, we recruited a sizable sample online ($N = 300$), and elicited the repetition masked priming response to low-, mid-, and high-frequency words with the prime SOA set at 33 ms. In general, our results showed that the online data collection may be reliably used for priming elicitation, modulo an additional pre-processing step to screen for accidental and uncontrollable fluctuations of the prime duration.Building on the validating results of experiment 1, we then designed experiment 2 so to address both issues mentioned above. First, we constructed two word frequency conditions on the basis of the more reliable SUBTLEX$_{US}$ word frequency database. The low and high-frequency conditions were built so to ensure that all words therein were seemingly known by the average English native speaker. Second, we ran an extensive power simulation analysis which took into account a number of parameter: sample size, standard deviation, correlation between priming conditions (related vs. unrelated), and raw interaction effect size. Our results showed significant main and interaction effects, with the masked priming response to low-frequency words being numerically about twice the masked priming response to high-frequency words.

Taken together, the results of the experiments reported here have important methodological and theoretical implications for the future research on priming and, more generally, visual word recognition. The methodological implications come from the extensive analysis we performed on the data we collected online (especially for experiment 1). We have shown that online data collection may provide reliable data from very large sample sizes–in the tens of thousands of subjects–something that could not even be imagined until a few years ago. Large-scale experimentation also provides a fast and effective solution to the long-standing issue of low statistical power–an issue that seem quite pervasive especially in psychology and related field. Of course, a potentially limitless source of participants entails a substantial increase of funding costs. These costs may sky-rocket especially for experimental designs such masked priming, in which time precision and accuracy are essential. As current online programs are not able to provide the timing standards required, the only feasible solution for the time being is to increase even more the sample size, which will necessarily entail additional costs to cover. However, we anticipate, as technology progresses towards more and more powerful machines, a gradual resolution of the limitations in terms of timing performance and reliability of online programs, thus boosting comparability of the online collected data to the in-lab collected data even more.

The theoretical implications of the present study come from the repetition masked priming results of experiment 2, which we ensured it had a sample size large enough to reliably detect a potential interaction effect between priming and frequency (i.e., frequency attenuation effects), while estimating the size of such effect between 10 and 15 ms (i.e., about half of the prime SOA and, as a consequence, of the maximum size of the repetition priming response). The significant frequency attenuation effects we found challenge one of the main predictions of the entry-opening model, which has since been the most accepted model of word recognition. According to this model, though detectable only after target recognition, masked priming actual taps onto the early stage of lexical processing, and is therefore devoid of any potential influence of episodic memory. When the visual stimulus is presented, lexical entries are assigned to specific bins based on orthographic similarity. In the first step (fast search), a fast search, frequency-ordered procedure goes through the entries within a given bin and compares each of them with the the input stimulus, thus assigning to each entry a goodness of fit score. This comparison is very fast and crude, and sorts entries into (a) perfect (i.e., no difference is detected between the input and the entry), (b) close (i.e., small differences are detected), and (c) irrelevant matches (ie.e., substantial difference are detected). Any entry of type (a) or (b) is opened, so that the entry can be further analyzed and compared it to the input in the second step (verification). In this model, priming therefore arises when the lexical entry of the target is opened at the prime presentation, and can be seen as a savings effect. In the masked environment, the short duration of the prime prevents the evaluation stage from starting, thus potentially eliciting effects even for non-perfect match cases (e.g., orthographic or morphological priming). Crucially, this model assumes that priming contributes to the entry opening procedure of the target before the target is actually presented. Therefore, priming arises only *after* the correct entry for the target has been located, and cannot be affected by frequency (Forster, Mohan, and Hector 2003, 17f.). Instead, the size of the masked priming response to low-frequency words reached about the prime SOA (33 ms), and therefore arguably maximal (as predicted by Forster 1999). More importantly, it doubled the size of the masked priming response to high-frequency words. Our statistical results confirm highly significant main effects of frequency (high vs. low) and primetype (unrelated vs. related); and significant interaction effects. This important result casts doubt on the supposed dissociation between word frequency and priming, and suggests that priming may actually take place even before the target entry has been located. Such an effect may arise in the fast search stage, assuming that the frequency-ordered search path takes longer for low-frequency words than for high-frequency words. When the entry that perfectly matched a low-frequency input words is opened, the priming effect will therefore be greater for these words than for high-frequency input words. Crucially, the modified model proposed above maintains the prediction regarding the size of priming as a function of prime duration (modulo factoring out any additional component potentially triggering hyper-priming effects: see Forster, Mohan, and Hector 2003 for further discussion). The experiments above seem to robustly corroborates such a prediction, as the size of the priming effects was numerically never above 30 ms (which roughly corresponds to the preset prime duration).

Setting aside any potential modification(s) of any given model accordingly, We ought to point out that these results offer at least two big-picture theoretical outcomes. Firstly, as these results may be factually accounted for, in one way or another, by all of three models considered here, they fail to help identification of the specific mechanisms underpinning word recognition and priming. Secondly, we anticipate the consequences of these results may reverberate in the related areas of research–in particular, morphological processing–that have primarily made use of the masked priming design for their investigations because of its supposed ability to tap onto the subliminal, and therefore strategy-free response. Studies on morphological processing have extensively made use of pairs in which the stem is primed as a stand-alone bare word (at least in languages where this is possible, e.g. English: *driver-DRIVE*). In the light of the results reported in this study, the size of the priming effects to such pairs should be affected by the frequency of the stem in a similar direction: i.e., low-frequency stems (i.e., *adornment-ADORN*) should trigger greater priming than high-frequency stems (i.e., *limitless-LIMIT*). If these predictions were to be true, they would further confirm that the supposed divide between early stages (involving automatic and "pre-lexical" processing) and later stages (involving instead access to more abstract information such as word frequency and semantic information) of word processing may not be as clear-cut as believed until now.

References

Angele, Bernhard, Ana Baciero, Pablo Gómez, and Manuel Perea. 2022. "Does Online Masked Priming Pass the Test? The Effects of Prime Exposure Duration on Masked Identity Priming." *Behavior Research Methods* 55 (1): 151–67. https://doi.org/10.3758/s13428-021-01742-y.

Balota, David A., Melvin J. Yap, Keith A. Hutchison, Michael J. Cortese, Brett Kessler, Bjorn Loftis, James H. Neely, Douglas L. Nelson, Greg B. Simpson, and Rebecca Treiman. 2007. "The English Lexicon Project." *Behavior Research Methods* 39 (3): 445–59. https://doi.org/10.3758/bf03193014.

Bodner, Glen E., and Michael E. J. Masson. 1997. "Masked Repetition Priming of Words and Nonwords: Evidence for a Nonlexical Basis for Priming." *Journal of Memory and Language* 37 (2): 268–93. https://doi.org/10.1006/jmla.1996.2507.

———. 2001. "Prime Validity Affects Masked Repetition Priming: Evidence for an Episodic Resource Account of Priming." *Journal of Memory and Language* 45 (4): 616–47. https://doi.org/10.1006/jmla.2001.2791.

Brysbaert, Marc, and Boris New. 2009. "Moving Beyond Kučera and Francis: A Critical Evaluation of Current Word Frequency Norms and the Introduction of a New and Improved Word Frequency Measure for American English." *Behavior Research Methods* 41 (4): 977–90. https://doi.org/10.3758/brm.41.4.977.

Brysbaert, Marc, and Michaël Stevens. 2018. "Power Analysis and Effect Size in Mixed Effects Models: A Tutorial." *Journal of Cognition* 1 (1). https://doi.org/10.5334/joc.10.

Cohen, Jacob. 1992. "A Power Primer." *Psychological Bulletin* 112 (1): 155–59. https://doi.org/10.1037/0033-2909.112.1.155.

Coltheart, Max, Kathleen Rastle, Conrad Perry, Robyn Langdon, and Johannes Ziegler. 2001. "DRC: A Dual Route Cascaded Model of Visual Word Recognition and Reading Aloud." *Psychological Review* 108 (1): 204–56.

Evett, Lindsay J., and Glyn W. Humphreys. 1981. "The Use of Abstract Graphemic Information in Lexical Access." *The Quarterly Journal of Experimental Psychology* 33 (4): 325–50.

Finger, Holger, Caspar Goeke, Dorena Diekamp, Kai Standvoß, and Peter König. 2017. "LabVanced: A Unified JavaScript Framework for Online Studies." In *2017 International Conference on Computational Social Science*. Cologne, Germany.

Forster, Kenneth I. 1998. "The Pros and Cons of Masked Priming." *Journal of Psycholinguistic Research* 27 (2): 203–33.

———. 1999. "Microgenesis of Priming Effects in Lexical Access." *Brain and Language* 68: 5–15.

Forster, Kenneth I., and Christopher Davis. 1984. "Repetition Priming and Frequency Attenuation in Lexical Access." *Journal of Experimental Psychology: Learning, Memory, and Cognition* 10 (4): 680.

———. 1991. "The Density Constraint on Form-Priming in the Naming Task: Interference Effects from a Masked Prime." *Journal of Memory and Language* 30 (1): 1–25. https://doi.org/10.1016/0749-596x(91)90008-8.

Forster, Kenneth I., Christopher Davis, C. Schoknecht, and R. Carter. 1987. "Masked Priming with Graphemically Related Forms: Repetition or Partial Activation?" *The Quarterly Journal of Experimental Psychology Section A* 39 (2): 211–51. https://doi.org/10.1080/14640748708401785.

Forster, Kenneth I., Kathleen Mohan, and Jo Hector. 2003. "The Mechanics of Masked Priming." In *Masked Priming: The State of the Art*, edited by Sachiko Kinoshita and Stephen J. Lupker, 3–37. New York, NY/Hove, UK: Psychology Press.

Grainger, Jonathan, and Arthur M. Jacobs. 1996. "Orthographic Processing in Visual Word Recognition: A Multiple Read-Out Model." *Psychological Review* 103 (3): 518–65.

Green, D. M., and J. A. Swets. 1966. *Signal Detection Theory and Psychophysics*. New York, NY: Wiley.

Kinoshita, Sachiko. 2006. "Additive and Interactive Effects of Word Frequency and Masked Repetition in the Lexical Decision Task." *Psychonomic Bulletin & Review* 13 (4): 668–73. https://doi.org/10.3758/bf03193979.

Kučera, J., and W. N. Francis. 1967. *Computational Analysis of Present Day American English*. Providence, RI: Brown University Press.

Leeuw, Joshua R. de. 2014. "jsPsych: A JavaScript Library for Creating Behavioral Experiments in a Web Browser." *Behavior Research Methods* 47 (1): 1–12. https://doi.org/10.3758/s13428-014-0458-y.

Lund, Kevin, and Curt Burgess. 1996. "Producing High-Dimensional Semantic Spaces from Lexical Co-Occurrence." *Behavior Research Methods, Instruments, & Computers* 28 (2): 203–8.

Masson, Michael EJ, and Glen E Bodner. 2003. "A Retrospective View of Masked Priming: Toward a Unified Account of Masked and Long-Term Repetition Priming." In *Masked Priming: The State of the Art*, edited by Sachiko Kinoshita and Stephen J. Lupker, 3–37. New York, NY/Hove, UK: Psychology Press.

McClelland, James L., and David E. Rumelhart. 1981. "An Interactive Activation Model of Context Effects in Letter Perception: Part i. An Account of Basic Findings."

Psychological Review 88 (5): 375–407.

Nievas, Francisco. 2010. "The Frequency Attenuation Effect in Identity and Associative Priming." *The Spanish Journal of Psychology* 13 (1): 30–62. https://doi.org/10.1017/s1138741600003668.

Norris, Dennis, and Sachiko Kinoshita. 2008. "Perception as Evidence Accumulation and Bayesian Inference: Insights from Masked Priming." *Journal of Experimental Psychology: General* 137 (3): 434–55. https://doi.org/10.1037/a0012799.

Peirce, Jonathan, Jeremy R. Gray, Sol Simpson, Michael MacAskill, Richard Höchenberger, Hiroyuki Sogo, Erik Kastman, and Jonas Kristoffer Lindeløv. 2019. "PsychoPy2: Experiments in Behavior Made Easy." *Behavior Research Methods* 51 (1): 195–203.

Petrosino, Roberto. 2020. "More Than Islands of Regularity: An Investigation of the Sensitivity of Morphological Decomposition to Higher-Level Linguistic Properties." PhD thesis, University of Connecticut.

Petrosino, Roberto, Jon Sprouse, and Diogo Almeida. submitted. "Asymmetries in the Stem and Suffix Masked Priming Response in a Large-Scale Online Study." In *Quaderni Di Linguistica e Studi Orientali*.

Potvin, Patrick J., and Robert W. Schutz. 2000. "Statistical Power for the Two-Factor Repeated Measures ANOVA." *Behavior Research Methods, Instruments, & Computers* 32 (2): 347–56. https://doi.org/10.3758/bf03207805.

Rajaram, Suparna, and James H Neely. 1992. "Dissociative Masked Repetition Priming and Word Frequency Effects in Lexical Decision and Episodic Recognition Tasks." *Journal of Memory and Language* 31 (2): 152–82. https://doi.org/10.1016/0749-596x(92)90009-m.

Scarborough, Don L., Charles Cortese, and Hollis S. Scarborough. 1977. "Frequency and Repetition Effects in Lexical Memory." *Journal of Experimental Psychology: Human Perception and Performance* 3 (1): 1–17. https://doi.org/10.1037/0096-1523.3.1.1.

Segui, Juan, and Jonathan Grainger. 1990. "Priming Word Recognition with Orthographic Neighbors: Effects of Relative Prime-Target Frequency." *Journal of Experimental Psychology: Human Perception and Performance* 16 (1): 65–76. https://doi.org/10.1037/0096-1523.16.1.65.

Sereno, Joan A. 1991. "Graphemic, Associative, and Syntactic Priming Effects at a Brief Stimulus Onset Asynchrony in Lexical Decision and Naming." *Journal of Experimental Psychology: Learning, Memory, and Cognition* 17 (3): 459–77. https://doi.org/10.1037/0278-7393.17.3.459.

Zeileis, Achim, and Gabor Grothendieck. 2005. "Zoo: S3 Infrastructure for Regular and Irregular Time Series." *Journal of Statistical Software* 14 (6): 1–27. https://doi.org/10.18637/jss.v014.i06.

Wordlists

Experiment 1

low frequency		mid frequency		high frequency	
unrel. word	word	unrel. word	word	unrel. word	word
levee	louse	smash	chasm	shoot	proof
clank	scone	manna	oxide	usual	clear
chafe	ardor	legit	vowel	teach	audio
paean	divvy	blunt	clerk	adult	apply
scamp	cacao	slope	bleed	allow	phone
shale	tulle	nasal	decor	forum	class
cilia	prole	forte	quirk	whole	raise
conga	neigh	aloud	speck	often	civil
burro	polyp	nymph	stash	issue	match
rummy	rebus	crass	ditch	style	local
chide	gipsy	squid	snare	coast	minor
scram	tibia	swirl	budge	reach	below
jaunt	licit	grunt	slack	smith	extra
hooey	sleet	taunt	sedan	speed	court
scald	qualm	cigar	tally	sense	exact
blare	swash	lunge	posit	write	bunch
whorl	bleep	negro	flock	trust	quick
trawl	kraut	exert	scorn	sleep	birth
sumac	duvet	lathe	grail	reply	truth
udder	trice	viola	bloat	track	serve
extol	nonce	rival	tumor	dream	trade
natty	fjord	dizzy	acute	image	heart
pleat	jetty	hertz	sauna	white	index
thrum	tipsy	haste	elect	flame	cable
groat	igloo	poppy	spoof	value	break
preen	hyena	clove	plush	avoid	woman
kazoo	frizz	guise	fiend	short	front
poach	bidet	magma	knelt	aware	voice
joule	therm	lotto	privy	large	stock
scull	piton	kayak	sigma	prove	seven
opine	phial	taint	parse	brand	blood
inure	scowl	fanny	carte	river	plain
acrid	scion	rouge	verge	guess	solid
droop	sinew	vitro	mourn	month	limit
dowse	prawn	floss	shrug	heard	scale
crepe	bleat	tempt	clasp	space	stuff
calve	kebab	flirt	bathe	leave	major
harpy	tawny	fluff	linen	agree	brown
cluck	exude	butch	stare	metal	house
beall	mosey	bowel	medic	along	stage
miser	skulk	aspen	weave	print	built
shuck	myrrh	chime	flint	worst	video

umber	schwa	crust	flank	sound	story
hooch	broil	spunk	scrub	faith	march
idyll	bijou	stoke	hoist	quote	clean
wrack	adorn	dairy	stout	train	price
douse	tiara	stale	cough	small	event
adieu	canny	gypsy	annex	night	thank
tepid	lolly	gloss	plume	shell	radio
joist	atoll	topaz	quart	alone	sorry

Experiment 2

low frequency		**high frequency**	
unrel. word	word	unrel. word	word
hobo	gull	half	shut
hark	glib	wait	dear
carp	acme	walk	like
watt	gala	rest	next
lush	wasp	word	girl
jive	pang	sure	town
burp	veer	call	case
sash	snug	miss	nice
hone	flux	high	stop
purr	snag	hope	wish
wham	dart	kill	side
brow	isle	lose	free
veto	bulk	turn	open
hoof	ruse	hate	fine
pant	yawn	give	tell
toot	toil	shot	make
rein	mitt	wife	play
acne	gout	hear	idea
heed	rift	pick	name
feud	spam	line	sick
cuss	meek	glad	part
mime	aura	move	deal
lore	zeal	hell	mean
afro	pear	body	door
ogre	rump	care	hurt
bard	vary	real	come
glee	cram	room	help
tuba	acre	late	soon
weld	muse	hard	fire
trio	malt	okay	even
glum	zinc	face	find
bead	gent	live	hold
daft	brim	show	time
sass	pulp	four	last
scab	duet	long	love

buoy	balm	mind	need
rant	spec	keep	know
slop	tram	down	talk
stow	haze	kind	five
mart	rune	year	home
lewd	skim	work	back
dill	posh	cool	baby
fizz	germ	send	hand
foil	prod	good	take
kale	kiwi	left	same
knit	lard	lady	stay
zest	flex	want	head
wail	halo	best	week
moan	jest	feel	meet
kilo	stub	city	read