

# *Pupil-BLAH-metry: Cognitive effort in speech planning reflected by pupil dilation*

**Megan H. Papesh & Stephen D. Goldinger**

**Attention, Perception, & Psychophysics**

ISSN 1943-3921  
Volume 74  
Number 4

Atten Percept Psychophys (2012)  
74:754-765  
DOI 10.3758/s13414-011-0263-y

## **Attention, Perception, & Psychophysics**

VOLUME 72, NUMBER 4 ■ MAY 2010

# AP&P

**EDITOR**

Jeremy M. Wolfe, *Brigham and Women's Hospital  
and Harvard Medical School*

**ASSOCIATE EDITORS**

Charles Chubb, *University of California, Irvine*

Bradley S. Gibson, *University of Notre Dame*

Simon Grondin, *Université Laval*

Lynne Nygaard, *Emory University*

Adriane E. Seiffert, *Vanderbilt University*

Joshua A. Solomon, *City University, London*

Shaun P. Vecera, *University of Iowa*

Yaffa Yeshurun, *University of Haifa*

A PSYCHONOMIC SOCIETY PUBLICATION

[www.psychonomic.org](http://www.psychonomic.org)

ISSN 1943-3921



**Your article is protected by copyright and all rights are held exclusively by Psychonomic Society, Inc.. This e-offprint is for personal use only and shall not be self-archived in electronic repositories. If you wish to self-archive your work, please use the accepted author's version for posting to your own website or your institution's repository. You may further deposit the accepted author's version on a funder's repository at a funder's request, provided it is not made publicly available until 12 months after publication.**

# Pupil-BLAH-metry: Cognitive effort in speech planning reflected by pupil dilation

Megan H. Papesh · Stephen D. Goldinger

Published online: 10 January 2012  
© Psychonomic Society, Inc. 2012

**Abstract** In reading research, a longstanding question is whether any stages of lexical processing require central attention, and whether such potential demands are frequency-sensitive. In the present study, we examined the allocation of cognitive effort in lexical processing by examining pupil dilations and naming latencies in a modified delayed naming procedure. In this *dual-task/change* procedure, participants read words and waited for various delays before being signaled to issue a response. On most trials (80%), participants issued a standard naming response. On the remaining trials, they were cued to abandon the original speech plan, saying “blah” instead, thereby equating production across different words. Using feature-matched low- and high-frequency words, we observed the differences in pupil dilations as a function of word frequency. Indeed, frequency-sensitive cognitive demands were seen in word processing, even after naming responses were issued. The results suggest that word perception and/or speech planning requires the frequency-sensitive allocation of cognitive resources.

**Keywords** Word recognition · Cognitive control · Automaticity · Attention

Word frequency effects are classic and robust in psycholinguistic research, predicting performance across many tasks, including lexical decision (Rubenstein, Garfield, & Millikan, 1970; Stone & Van Orden, 1993), naming (Forster & Chambers, 1973; Waters & Seidenberg, 1985), and perceptual identification (Manelis, 1977). Even in silent reading, word frequency has powerful effects on eye

movements and fixation durations (Inhoff & Rayner, 1986; Rayner & Duffy, 1986). Despite the ubiquity of frequency effects, questions remain about their locus in word processing: Are frequency effects restricted to early processes of perception and lexical access, or do they continue into postaccess processes? A related question concerns the automaticity of lexical access: Does word perception require central attention, and might cognitive demands differ across high-frequency (HF) and low-frequency (LF) words?

The word frequency effect is most commonly assessed using response times (RTs); across variations in methods, stimuli, and participants, people are faster to process HF words, relative to LF words. Critically, however, frequency interacts with many other variables, such as context (Becker & Killion, 1977), stimulus quality (Yap, Balota, Tse, & Besner, 2008), word repetition (Scarborough, Cortese, & Scarborough, 1977), and neighborhood density (Andrews, 1992; Grainger & Segui, 1990). Frequency effects also change across tasks; for example, they are approximately three times larger in lexical decision, relative to naming. As suggested by Balota and Chumbley (1984), this interaction suggests that the decision stage (a necessary component for lexical decision, but not for naming) is frequency-sensitive. Thus, frequency appears to affect processes beyond perception in lexical decision. Indeed, Balota and Abrams (1995) later reported that frequency influences response mechanics (e.g., arm movement force) in lexical decision. In contrast, the naming task does not entail a decision stage, and is often assumed to provide a “cleaner” estimate of frequency (and other lexical) effects.

## The delayed naming task

In a standard naming task, participants try to correctly name stimulus words as quickly as possible following their

M. H. Papesh · S. D. Goldinger (✉)  
Department of Psychology, Arizona State University,  
Box 871104, Tempe, AZ 85287-1104, USA  
e-mail: goldinger@asu.edu

presentation (e.g., Forster & Chambers, 1973). Given this procedure, however, an RT difference between LF and HF words might reflect lexical access, or it might reflect uncontrolled differences in articulatory complexity (or other factors) across the words. To address this issue, the *delayed naming task* was developed. In delayed naming, the participant sees a word and then waits, often for variable intervals, for a “go” signal to initiate the naming response. The logic is simple: Because lexical access is allowed to finish, any remaining frequency (or other) effect should arise from speech planning or execution. As reviewed by Goldinger, Azuma, Abramson, and Jain (1997), frequency effects have been inconsistently reported in delayed naming. Indeed, the procedure was originally used as a control task, to ensure that frequency effects in naming times did not unduly reflect phonetic variations across items (Andrews, 1989; Forster & Chambers, 1973). Conversely, in a well-known study, Balota and Chumbley (1985) had participants view HF and LF words, wait for response cues at various delays, and then quickly speak. Balota and Chumbley observed reliable frequency effects following relatively long delays (650 ms or longer), suggesting that frequency affected post-access processes. These findings have been replicated and extended (Connine, Mullennix, Shernoff, & Yelen, 1990), but have also been questioned (McRae, Jared, & Seidenberg, 1990). For example, whenever participants must produce different words, there remains a potential for uncontrolled phonetic variations, leading to differences in voice-key activation.

To address the issue of interword phonetic variations, Goldinger et al. (1997) modified the delayed naming task not only to wait out lexical access, but also to equate production, leaving only the speech-planning process free to vary. To accomplish this purpose, they used a *dual task/change* procedure. Most trials (80%) entailed “standard” delayed naming, in which participants saw words, waited for variable delays, and spoke the words in response to a high-pitched tone. In the remaining, catch trials, a low-pitched tone indicated that participants should abandon the speech plan, quickly saying “blah” instead. Using delays ranging from 150 to 1,400 ms, Goldinger et al. (1997) observed robust frequency effects in both standard and catch trials, with slower responses following LF words. That is, people were slower to produce LF words, as well as slower to switch away from LF words to the “blah” response. The results suggested that word frequency affects speech planning, independent of speech production, with LF words demanding extra attention. Still, other researchers have employed the delayed naming procedure and have not observed any reliable frequency effects (McCann & Besner, 1987; McRae et al., 1990; Savage, Bradley, & Forster, 1990). In general, frequency effects in delayed naming are fragile, appearing and disappearing with changes to the

stimuli and methods. As such, it is difficult to conclude that word frequency truly affects lexical processing beyond the early stages of perception.

### Pupillometry

Given such fragile effects in delayed naming, one goal of the present study was to widen our focus, expanding from relying on individual RTs to examining a more continuous estimate of cognitive effort. Our specific approach was to collect continuous estimates of pupil dilation while participants performed a modified delayed naming task, as in Goldinger et al. (1997). This approach allowed for the assessment of potential frequency effects in delayed naming, independent of the presence or absence of an RT effect. Although the pupils change reflexively in response to general factors, such as emotional arousal, stress, and anxiety, such *tonic* changes are independent of *phasic* changes, which are based on the onset of stimuli for cognitive processing (Karatekin, Couperus, & Marcus, 2004). Phasic changes in pupil diameter have been used to infer cognitive effort across many domains, including lexical decision (Kuchinke, Võ, Hofmann, & Jacobs, 2007), attention allocation (Karatekin et al., 2004), working memory load (Granholm, Asarnow, Sarkin, & Dykes, 1996; Van Gerven, Paas, Van Merriënboer, & Schmidt, 2004), face perception (Goldinger, He, & Papesch, 2009), and general cognitive processing (Granholm & Verney, 2004).

The appeal of pupillometry for the investigation of cognition lies in its relative independence: Pupil dilation and constriction occur automatically, so cognitive effort can be gauged without influences from task-specific strategies. Prior research has shown that, when people perform more difficult cognitive operations, their pupils dilate (Porter, Troscianko, & Gilchrist, 2007), which may represent the summed index of the required brain activity (Beatty, 1982). Task-evoked peak pupil diameters have been used to index the cognitive load associated with memory tasks (Kahneman & Beatty, 1966), showing that pupils dilate relatively quickly with the onset of mental effort and constrict relatively quickly after its cessation. Indeed, Kahneman (1973) used the pupillary reflex as his primary index of mental processing load in his theory of attention, citing its sensitivity to variations within or between tasks and its ability to reflect individual differences in cognitive ability. Relevant to the present study, Kuchinke et al. (2007) recorded pupil dilation during a lexical decision task, using HF and LF words of varying emotional valences. Average peak pupil dilation in response to LF words significantly exceeded peak dilation to HF words (approximately 1,200 ms after stimulus onset), suggesting that pupil dilation can indirectly reflect the cognitive effort devoted to word processing. Given their use of

lexical decision, however, the role of the decision stage remains unclear. A central aim of the present study was to assess “hidden” frequency effects in lexical processing without requiring lexical decisions.

When using pupillometry, particular caution must be exercised in stimulus generation and development of the procedural details. The pupils dilate reflexively to changes in the luminance, color, and spatial frequencies of visual input. Therefore, care must be taken to equate, as closely as possible, the stimulus characteristics across conditions. Porter and Troscianko (2003) identified several methods to minimize unwanted pupillary reflexes, including the use of relatively low stimulus contrast, avoiding colored stimuli, and using relatively long stimulus exposure durations. To address these considerations in the present investigation, we maintained constant background lighting and a constant screen color and (of greatest importance) equated the visual features of our stimulus words as closely as possible. Although we presented words briefly (500 ms), we examined extended time courses for pupillary reflexes, by utilizing response tone delays up to 2,000 ms and extending trials up to 6,000 ms postresponse. Examining an extended time course is critical, because pupillary reflexes appear several hundred milliseconds after the onset of stimulus processing (see Kuchinke et al., 2007). It is important to note, however, that pupillary responses to cognitive activity occur within approximately 250 ms (Beatty, 1982) and can be used to examine activities (such as reading; Heitz, Schrock, Payne, & Engle, 2008) that extend continuously in time.

In the present research, we adopted the modified, dual-task delayed naming paradigm used by Goldinger et al. (1997) and continuously monitored participants' pupil diameters to estimate cognitive effort during speech planning. After reading briefly presented words, the participants waited for a response tone to indicate one of two responses, with different tones signaling either a standard naming response or a catch response (i.e., “blah”).<sup>1</sup> Response tones occurred following one of four randomly selected delays, ranging from relatively short (250 ms) to relatively long (2,000 ms). Given the nature of phasic pupillary reflexes, we expected to observe diverging pupil diameters for LF and HF words across longer delays. Specifically, if word frequency influences speech planning, preparing to speak LF words should demand greater cognitive resources, reflected by relatively enlarged pupils, following the offset of the printed word. Furthermore, this result should persist even when participants eventually execute a catch response, in which HF and LF words are phonetically equal. That is,

<sup>1</sup> Note that the present method is agnostic with respect to cascaded versus discrete stages of processing. By allowing perceptual processes to complete and equating production (during catch trials), our method allowed us to examine the influence of word frequency during speech planning.

we expected pupil dilation to reflect the greater effort required to abandon the speech plans for LF words. Although results in delayed naming RTs have been equivocal, we expected to observe a physiological manifestation of frequency effects by examining pupil diameters across standard and catch trials. In theoretical terms, such a finding would indicate that speech planning requires central attention, and that its demands vary across words in a frequency-sensitive manner. To foreshadow our results, we observed a frequency effect in standard delayed-naming RTs, but not in catch-trial RTs. However, we did observe clear differences in the allocation of cognitive effort across HF and LF words, which persisted for several hundred milliseconds after naming responses.

## Method

### Participants

A group of 43 native English speakers (18–30 years of age,  $M = 19.50$ ,  $SD = 2.2$ ) from Arizona State University participated for partial course credit. All of the participants reported normal or corrected-to-normal vision and no hearing difficulty. Two of the participants were dropped from the analysis because of missing pupil data (greater than 7% of their data).

### Stimuli

We compiled a list of 150 word pairs, matched for visual features, length, and syllables, but differing in rated word frequency (see Table 1 and Appendix A). To match the visual features, we selected word pairs that differed by one letter (e.g., *few/pew*) or the order of letters (e.g., *great/grate*), selecting homophones whenever possible. The LF and HF word sets were closely matched for phonetic onsets, with a slight bias in favor of LF words (i.e., more initial stops). The HF words featured 88 initial stops, 43 initial fricatives, and 19 initial liquids/glides, and the LF words featured 92 initial stops, 38 initial fricatives, and 19 initial liquids/glides. The words were pseudorandomly assigned to five lists, such that each list contained 60 words (48 standard naming and 12 catch), which were presented to participants in random order. Across participants, all words were used in catch trials equally often. Response tones were generated using NCH Tone Generation software. Of the participants, 22 discriminated between tones of 900 and 400 Hz, and the other 21 discriminated between tones of 750 and 550 Hz.<sup>2</sup>

<sup>2</sup> Although we expected the difficulty of the response tone discrimination to interact with any observed frequency effects, this was not borne out by the data. All of the analyses were collapsed across differences in the response tones.

## Procedure

Participants were tested individually in a dimly lit, sound-attenuated booth. A chinrest maintained head position and viewing distance at 60 cm. The stimuli were presented in lowercase, 24-pt black Arial font on a constant gray background (RGB 150) at 1,024 × 768 resolution on a 17-in. monitor. Eye movements and pupil diameter were continuously recorded binocularly at 50 Hz, using a Tobii 1750 eyetracker. Naming latencies were recorded by an SR response box with a voice key. The experiment was controlled and data were collected using E-Prime 1.1 software (Psychology Software Tools, 2006).

Participants were first familiarized with the experiment and the eyetracker. The chinrest was adjusted such that eye position was maintained centrally on the horizontal axis and slightly above center on the vertical axis, and then participants were calibrated. For the calibration routine, we randomly presented nine fixation points (indicated by a blue dot) over the range of the display; participants “followed the dot” as it moved to each location. If the software or researcher identified any missing fixations, the calibration routine was repeated. All of the participants were successfully calibrated within two attempts.

The experiment began with a tone identification task, wherein participants judged the response tones as being either high- or low-pitched by pressing a corresponding response key. All of the participants completed 6 trials with 100% accuracy. After reading the instructions and having the task verbally explained, participants completed 6 practice trials (half standard and half catch). If the participant committed an error on any trial, the researcher reminded them of the instructions before proceeding with the experimental trials. Each experimental trial proceeded as outlined in Fig. 1, with 240 standard naming trials and 60 catch trials. Following the offset of a 1,500-ms fixation cross, participants were shown a single word for 500 ms. After the offset of the word, participants waited over one of four delay periods (250, 500, 1,000, or 2,000 ms) before a pseudorandomly selected high- or low-pitched tone indicated that a verbal response should be made. High-pitched tones signaled standard naming trials; low-pitched tones signaled catch trials. The tones were selected to maintain the ratio of standard to catch trials used by Goldinger et al. (1997); only 20% of the trials were catch trials, a percentage meant to encourage standard speech planning during the delay period. Following their verbal responses, participants were given at least 4,000 ms (but up to 6,000 ms) for their pupils to constrict to a preestablished baseline range.<sup>3</sup> Short breaks were permitted after every 60 trials.

<sup>3</sup> In pilot tests without a 6,000-ms blank screen, we found that frequency strongly influenced pupil diameters in subsequent trials. Additional pilot testing identified 6,000 ms as the maximum time necessary for the pupils to sufficiently constrict in order to eliminate artifacts from preceding trials.

**Table 1** Summary statistics for the stimulus set

	Low-Frequency Words	High-Frequency Words
Mean frequency (KF) †*	6.67 (6.5)	193.18 (343.5)
Mean frequency (HAL)*	3,047.36 (3,486.1)	90,698.60 (140,896.6)
Mean frequency (Subtitle) ††*	8.16 (17.7)	208.74 (402.9)
Mean orthographic $N$	7.11 (5.2)	7.35 (5.4)
Mean orthographic $N_f$ *	8.74 (1.1)	8.20 (1.3)
Mean phonologic $N$	16.08 (11.4)	16.35 (11.4)
Mean phonologic $N_f$	8.18 (1.1)	8.16 (0.9)
Mean length (letters)	4.55 (0.8)	4.55 (0.8)
Mean length (phonemes)	3.62 (0.8)	3.56 (0.8)

$N$  = neighborhood.  $N_f$  = neighborhood frequency. All data were obtained from Balota et al. (2007). The data in parentheses are standard deviations. † From Kučera and Francis (1967). †† From Brysbaert and New (2009). \*  $p < .05$

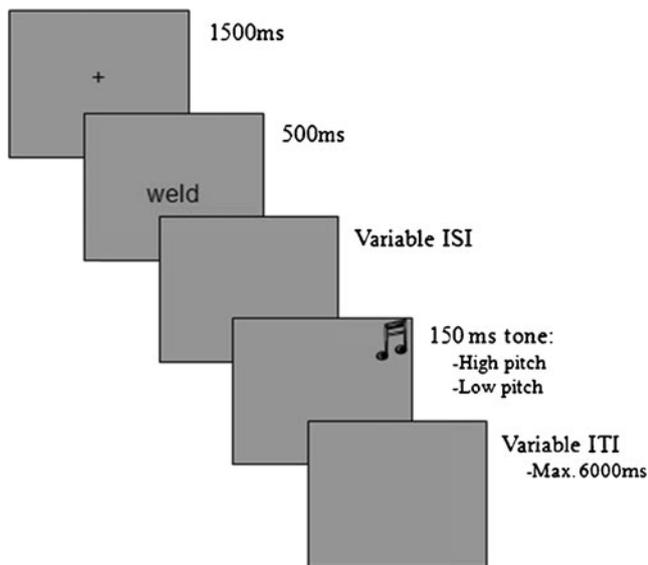
## Results

Trials with errors (i.e., mispronunciations, pretone responses, and voice key errors) were removed from the analysis.<sup>4</sup> Missing pupil data (due to blinks) were filled in by linear interpolation, and peak diameters were averaged during specific trial events (fixation, word presentation, wait, tone, response preparation, response, and postresponse intertrial interval [ITI]). The missing pupil data were not systematically distributed across experimental variables, and none of the 41 participants whose data were retained had more than 6% missing data. Baseline-corrected peak pupil diameters were computed, on a trial-by-trial basis, as the difference between the average of the participant's pupil diameter during the fixation cross and the event peak of the current trial. This procedure minimized the influence of “carryover” effects from the word frequency or difficulty of the preceding trial. All pupil analyses were conducted on baseline-corrected data from each participant's right eye. For all analyses, alpha was maintained at .05, and multiple comparisons were corrected with Bonferroni adjustments.

### Naming latency

A total of 5 participants with excessive naming errors (greater than 10% of trials) were excluded on a case-wise basis,

<sup>4</sup> Participants averaged 17.44 errors each. Of the 715 total errors committed, 502 were voice-key errors (i.e., the participant did not speak loud enough to trigger the voice key), 147 were mispronunciations, 13 were pretone responses, 49 were catch trial errors, and 4 were standard trial errors (i.e., saying “blah” in a standard trial). Generally speaking, all errors were more likely when trials contained LF words, but we refrained from analyzing the data, given the preponderance of voice-key errors.



**Fig. 1** Schematic outline of a single experimental trial

leaving 36 in the analysis. The naming RTs for all trials are summarized in Table 2. As shown, we observed robust effects of delay (primarily in the catch trials), but the effects of word frequency were inconsistent. These data were first analyzed in an omnibus  $2$  (word frequency: HF or LF)  $\times$   $2$  (trial type: standard or catch)  $\times$   $4$  (delay: 250, 500, 1,000, or 2,000 ms) within-subjects repeated measures (RM) ANOVA. Although a main effect of delay,  $F(3, 33) = 55.18$ ,  $p < .001$ ,  $\eta_p^2 = .83$ , indicated that participants responded faster as the delay increased, an interaction between delay and trial type,  $F(3, 33) = 45.25$ ,  $p < .001$ ,  $\eta_p^2 = .80$ , revealed that this pattern was only evident during catch trials. Standard trials (784 ms) were faster than catch trials (1,169 ms). This main effect of trial type,  $F(1, 35) = 146.63$ ,  $p < .01$ ,  $\eta_p^2 = .81$ , reflected the difficulty associated with task switching. In the omnibus analysis, we did not observe a reliable effect of word frequency ( $F < 1$ ); the average naming latency for HF words (979 ms) was equivalent to that for LF words (974 ms).

Following the omnibus analysis, we conducted separate  $2 \times 4$  (Word Frequency  $\times$  Delay) RM ANOVAS on the standard trials and the catch trials. In the standard trials, we observed a reliable, 17-ms effect of word frequency,  $F(1, 35) = 7.48$ ,  $p < .01$ ,  $\eta_p^2 = .18$ . The main effect of delay was marginal,  $F(3, 105) = 2.46$ ,  $p = .07$ ,  $\eta_p^2 = .06$ , reflecting the small (20-ms) decrease in average naming RTs across delays. The interaction was not reliable ( $F < 1$ ). In the catch trials, the main effect of word frequency was not reliable ( $F < 1$ ), but there was a robust effect of delay,  $F(3, 105) = 67.92$ ,  $p < .01$ ,  $\eta_p^2 = .66$ , reflecting the large (964-ms) decrease in RTs across delays. The interaction was not reliable ( $F < 1$ ). Taken together, the naming latencies partially replicated the prior results from Goldinger et al. (1997), although two effects (delay in standard trials and

**Table 2** Average naming response times (in milliseconds) in all conditions

Tone Delay	Standard Trials		Catch Trials		Mean
	HF	LF	HF	LF	
250 ms	781 (24)	811 (25)	1,696 (81)	1,722 (84)	1,253
500 ms	784 (23)	791 (29)	1,325 (63)	1,291 (79)	1,048
1,000 ms	774 (22)	777 (22)	907 (68)	918 (54)	844
2,000 ms	762 (22)	791 (23)	799 (66)	691 (29)	761
Mean	775	793	1,182	1,156	

Standard errors are shown in parentheses

frequency in catch trials) were weaker than had been previously reported. We consider these results in the Discussion section, after examining the pupillary results.

### Pupil dilation

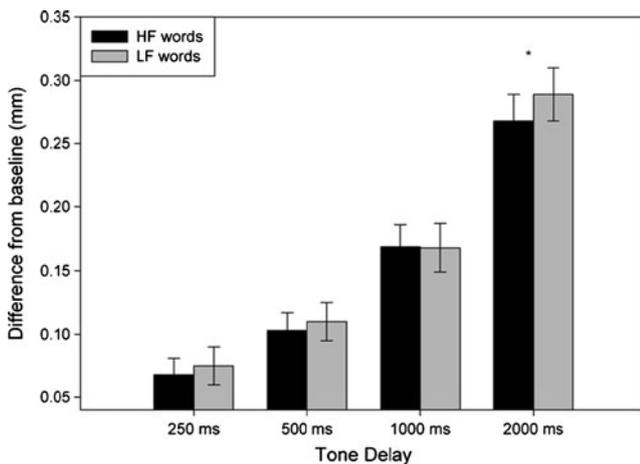
Baseline-corrected peak pupil diameters were analyzed in separate  $2$  (word frequency)  $\times$   $2$  (trial type)  $\times$   $4$  (delay) within-subjects RM ANOVAs for each of the seven trial events. That is, pupil diameters were not examined in veridical “real time,” but in time-invariant sequences of trial events. For example, the “postresponse” trial period in the 250-ms delay condition occurred in real time at approximately 2,100 ms; this was the real-time equivalent of the “wait” trial period in the 2,000-ms delay condition. Because of such timing differences, pupillary responses were analyzed according to trial events. (Note, however, that the same results were observed in a set of real-time analyses.) Complementary analyses were carried out for each delay condition separately, on baseline-corrected, event-locked peak pupil diameters. For ease of presentation, these results are discussed in order of the trial events, as peak millimeter differences from baseline ( $\text{mm}_d$ ). The full ANOVA results can be found in Appendix B; only results of theoretical interest are discussed in the text.

The relatively late-arriving characteristic of the pupillary response suggests that reliable effects should occur at least 200–300 ms after a relevant cognitive event (Beatty, 1982). Thus, we do not address analyses on the fixation and word trial events (indeed, there was no effect of word frequency while words were on screen,  $p = .47$ ). After word offset, participants waited for variable delay periods before hearing the response tone. During this wait period, we observed a main effect of trial type: Although participants had no way of knowing what trial was coming, their peak diameters were reliably larger during standard (0.17  $\text{mm}_d$ ) than during catch (0.14  $\text{mm}_d$ ) trials. We suspect that this effect is spurious, driven primarily by having relatively few catch trials. We also observed a strong main effect of delay during the wait period, and a marginal effect of word frequency. As is

shown in Fig. 2, as the delay period grew longer, participants' peak pupil diameters increased, with a general trend toward larger pupil diameters in trials with LF words.

After hearing the response tone, participants either initiated execution of the predominant naming response or switched to prepare (and execute) the alternative “blah” response. During this response preparation period, we observed main effects of delay and word frequency. As was evident during the wait period, longer delays produced greater peak dilation. Consistent with our predictions and previous findings (Kuchinke et al., 2007), the main effect of word frequency revealed that LF words elicited greater dilation ( $0.38 \text{ mm}_d$ ) than did HF words ( $0.34 \text{ mm}_d$ ). Consistent with the notion of cognitive effort, catch trials placed a greater demand on participants, relative to standard trials. Although this trend was consistent across all delays, the Delay  $\times$  Trial Type interaction indicated that the difference was only reliable in the 250- and 500-ms delay conditions: Preparing a “blah” response resulted in enlarged pupils at short delays. As is shown in Fig. 3, the difference between LF and HF words was numerically larger when participants planned a catch trial response, suggesting that switching away from the predominant speech plan was more resource-demanding for LF words ( $0.35 \text{ mm}_d$ ) than for HF words ( $0.33 \text{ mm}_d$ ).

When participants correctly issued verbal responses, we again observed several reliable patterns. There was a frequency effect: Issuing an LF response (either standard or catch) demanded greater effort ( $0.25 \text{ mm}_d$ ), relative to emitting an HF response ( $0.18 \text{ mm}_d$ ). Although it was present in both standard and catch trials, the frequency effect was stronger in catch trials ( $\eta_p^2 = .62$ ) than in standard trials ( $\eta_p^2 = .14$ ). Actually producing the catch response, however, demanded *fewer* resources ( $0.16 \text{ mm}_d$ ), relative to standard naming ( $0.27 \text{ mm}_d$ ). We interpret this finding as a speech-motor benefit for repeated items. On average, participants



**Fig. 2** Average baseline-corrected peak pupil diameters during the variable wait period, collapsed across trial types. Error bars represent standard errors. \* $p < .05$

successfully issued more than 50 “blah” responses per session. Over time, the associated motor program may have become well-practiced, demanding fewer resources.

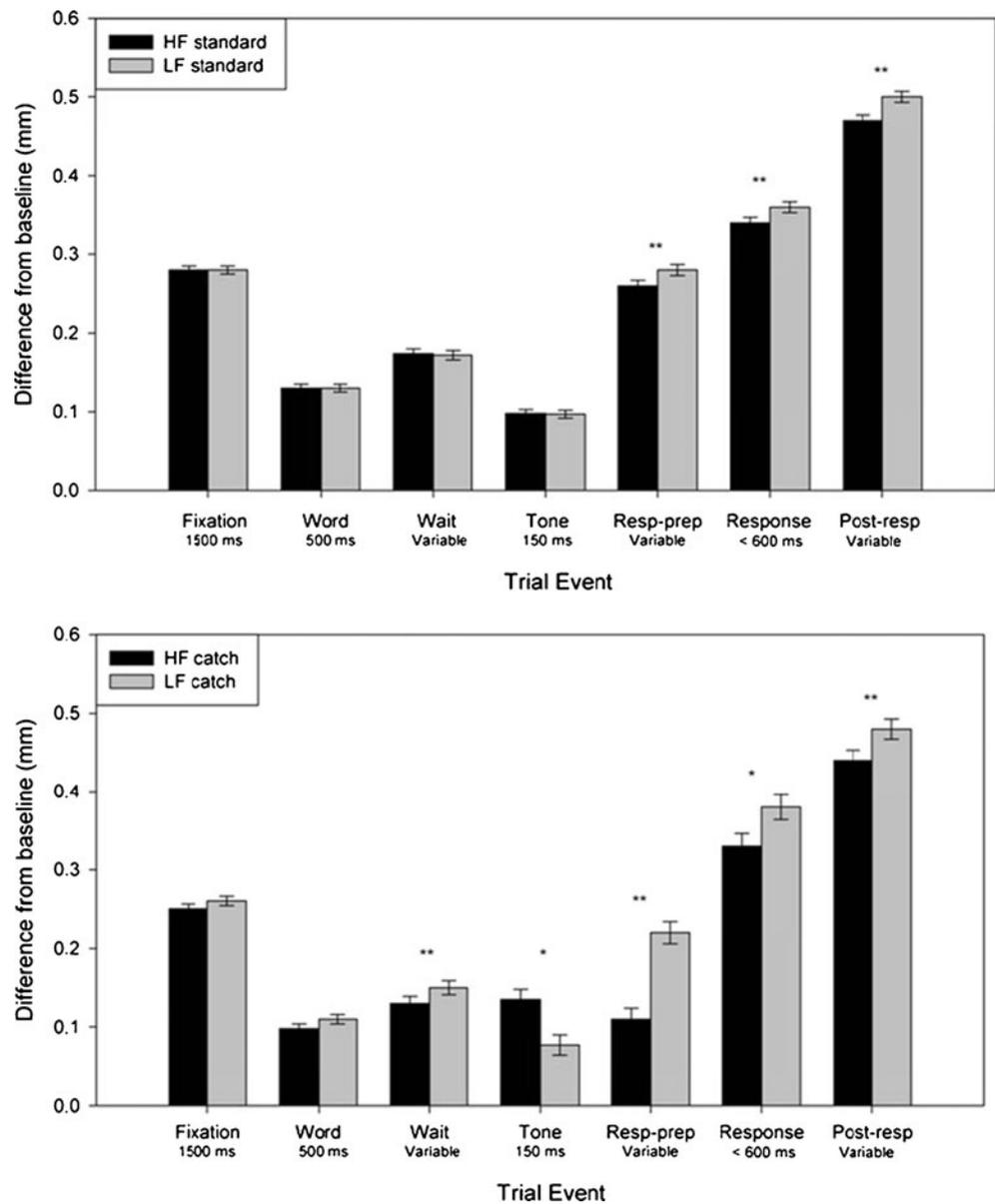
After participants issued their spoken responses, they waited for variable ITIs (up to 6,000 ms) before the start of the next trial. During the postresponse period, we again observed a reliable main effect of word frequency, suggesting that the physiological reflection of the effort devoted to LF words levels off slowly, relative to the HF words: Whereas pupil diameters following HF words began to quickly constrict during this period ( $0.45 \text{ mm}_d$ ), diameters following LF words were still more dilated ( $0.49 \text{ mm}_d$ ). The main effect of delay was again evident, but was only reliable in catch trials.

Considering this postresponse period, one aspect of the results in Fig. 3 requires further explanation. Specifically, it appears that pupil diameters increased in the postresponse period, relative to the moment when participants actually spoke. As was shown by Kahneman and Beatty (1966), pupils typically constrict immediately upon the cessation of a cognitive load. The apparent increase shown in Fig. 3 is actually an artifact of the naming procedure. In a naming task, RTs are recorded from the moment a voice key is triggered and are typically the only dependent measure. In the present study, we continued collecting pupil data throughout the spoken response, when effort was still being expended (movement also affects the pupil reflex; see Richer & Beatty, 1985). Because Fig. 3 represents peak dilation across trial events, each value represents the average maximum, rather than the true average. To better illustrate the underlying pupil waveforms, Fig. 4 shows average pupil diameters for all standard and catch trials in the 500-ms delay condition. For clarity, the data are shown as raw values in millimeters, not corrected for baseline differences. As the figure makes evident, average pupil diameters were generally higher for LF words throughout all trial periods, with the peak typically corresponding to the spoken response in standard trials, and to the “switch” decision in catch trials. The upward deflection following the tone onset was also sharper for catch relative to standard trials, indicative of task-switching effort. When they are viewed as direct averages, it is evident that pupil diameters indeed returned toward baseline once responses were issued.

## Discussion

In the present study, we assessed the relative cognitive demands of naming LF and HF words. By coupling a modified delayed naming procedure with pupillometry, we examined both standard (i.e., naming latencies) and novel (i.e., pupillary reflexes) manifestations of the word frequency effect. In the majority (80%) of experimental trials, participants issued standard delayed naming responses,

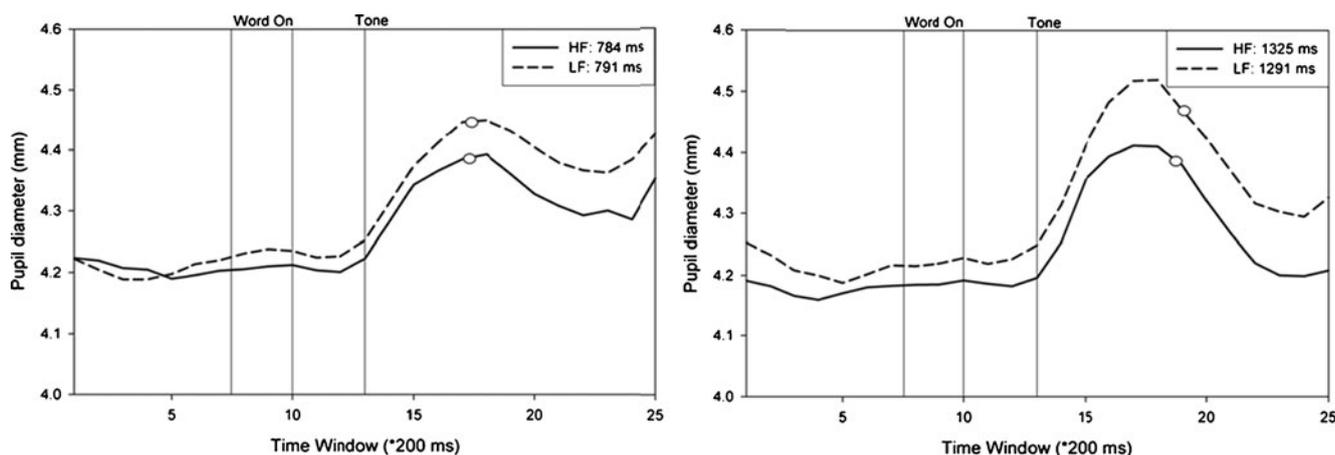
**Fig. 3** Baseline-corrected peak pupil diameters for high-frequency (HF) and low-frequency (LF) words during each trial period for standard (top panel) and catch (bottom panel) trials, collapsed across delay conditions. Error bars represent standard errors. \* $p < .05$ . \*\* $p < .01$



following the perception of HF or LF words. In the remaining trials, they had to abandon the speech plan, saying “blah” instead, thereby equating speech production across HF and LF words. Our results generally replicated and extended those from Goldinger et al. (1997), although some differences were observed. Considering first the similar findings, we verified that standard naming is faster than switching to the “blah” response. Second, across all delay periods, we observed slower standard naming latencies for LF than for HF words. Although we did not observe frequency effects in the catch trials, the pupillary results support and extend those from Goldinger et al. (1997): Using pupil dilation as a proxy for cognitive effort (Kahneman, 1973), we observed that preparing to speak LF words, relative to HF words, demanded greater cognitive resources, a difference that persisted even after speech programs had been

executed. The results suggest that lexical processes can be examined by monitoring pupillary reflexes (Kuchinke et al., 2007), potentially revealing differences in cognitive demands, even in cases with equivalent overt performance (Karatekin et al., 2004).

Although the pupillary results support earlier findings, it is important to note that two results from Goldinger et al. (1997) were not replicated in the present RT data. First, in the standard naming trials, the effect of delay was small. Whereas responses typically become faster with longer naming delays (Balota & Chumbley, 1985; Goldinger et al., 1997; Kawamoto, Liu, Mura, & Sanchez, 2008), we observed only a 20-ms effect. In contrast, the catch trials showed a robust delay effect. Together, these results suggest that participants may have been anticipating switch (“blah”) trials to an unreasonable degree. In a result of greater



**Fig. 4** Average pupil diameters (in millimeters) for high-frequency (HF) and low-frequency (LF) words from all trials with a 500-ms delay. The left panel shows results from standard trials, and the right panel shows catch trials. In both panels, the dots reflect average

naming response times, with values shown in the inset boxes. The trial periods corresponding to word presentation and tone onset are indicated by vertical lines

relevance to the present article, although we observed a reliable (albeit small) frequency effect in the standard delayed naming trials, we found no such effect in the catch trials, and thus failed to replicate this result from Goldinger et al. (1997).

We have two main hypotheses regarding the differences across the studies. Our primary hypothesis is that the present stimulus items were chosen in a manner that maximally preserved visual features across matched LF and HF words, but also minimized phonetic differences. Many pairs contained similar sound patterns (e.g., *drawn/drown*, *mouse/moose*, *sweet/sweep*), and some were fully homophonic (e.g., *course/coarse*, *pain/pane*, *read/reed*). We selected items with strict criteria in order to avoid unwanted pupillary reflexes, and we verified (in a pilot study) that they produced a reliable frequency effect in standard naming. Although the executed motor plan was always the same in catch trials, the initial speech motor plan (i.e., speaking the printed word) was left free to vary, although perhaps not enough to elicit a difference in task-switching time. Once word perception was complete in delayed naming, it may have been that the speech–motor programs were too similar across LF and HF words to reliably elicit a difference.

Our second hypothesis regards the pace of the experiment—in particular, its overall slow RTs. In the present experiment, the average RT in standard delayed naming was approximately 780 ms; analogous RTs in Goldinger et al.'s (1997) study were approximately 600 ms. Once again, this difference may reflect the phonetic composition of the stimulus materials. The present experiment included many items with “soft” initial phonemes, which are well known to affect voice-key responses (Kawamoto et al., 2008; Kessler,

Treiman, & Mullennix, 2002). It is also possible that participants responded slowly because of the pace of the experiment. In order to allow pupils adequate time to return to baseline dilation, we introduced long ITIs (from 4 to 6 s) that may have slowed performance. The present experiment lacked the rhythm of most naming experiments, which has been shown to strongly affect RTs (see Kello, 2004). Our procedure also involved a forced choice response (standard/catch) after the response tone, whereas most delayed naming procedures merely involve naming. By introducing an extra task demand, we may have disrupted the normal rhythm of delayed naming, thereby slowing naming latencies and obscuring RT-based frequency effects. Whether either hypothesis is correct, it is important to emphasize that RTs represented only one dependent measure. Pupil dilation, measured in those same trials, revealed effects that conceptually replicated those from Goldinger et al. (1997), suggesting that the lower-frequency words imposed greater cognitive load.<sup>5</sup>

As noted earlier, word frequency effects are ubiquitous in perceptual identification (Goldiamond & Hawkins, 1958; Manelis, 1977), lexical decision (Whaley, 1978), silent reading (Inhoff & Rayner, 1986; Rayner & Duffy, 1986), recognition memory (Glanzer & Adams, 1985), and other domains. Despite the theoretical importance of word frequency, debate surrounds its locus in the stream of cognitive processing. For example, although word frequency clearly affects perception (O'Malley, Reynolds, Stolz, & Besner,

<sup>5</sup> Note that the relationship between RTs and pupil dilation is complex, and the measures do not typically correlate with each other (see Richer & Beatty, 1985, 1987; Richer, Silverman, & Beatty, 1983).

2008; but see McCann & Besner, 1987; Paap & Johansen, 1994), interactions between word frequency and tasks (e.g., lexical decision vs. naming) suggest that the decision stage is also frequency-sensitive. To examine decision-free frequency effects, researchers may use the delayed naming procedure, wherein naming occurs once lexical access is complete. If word frequency exerts an influence solely during the perception and decision stages, it should disappear in delayed naming; however, if postaccess processes are frequency-sensitive, the differences should persist. We observed the latter result, suggesting that frequency modulates the cognitive effort required for speech planning (Goldinger et al., 1997; see also Herdman, 1992).

With the exception of event-related potential (Dambacher, Kliegle, Hofmann, & Jacobs, 2006; Hauk & Pulvermüller, 2004; Rugg, 1990) and fMRI (Chee, Westphal, Goh, Graham, & Song, 2003; Kronbichler et al., 2004) investigations, frequency effects are typically observed as overt behavioral differences across tasks, such as naming times or fixation durations. Although behavioral measures clearly identify frequency-based differences in lexical processing, they may mask dynamic attentional differences that persist after the offset of a behavioral response. In our study, we examined naming times but also estimated cognitive demands via pupillometry. As noted earlier, researchers have long examined the pupillary reflex to study cognitive processes, including lexical decision (Kuckinke et al., 2007), visual search (Porter et al., 2007), attention and concentration (Bradshaw, 1968), imagery (Paivio & Simpson, 1966), and memory (Papesh & Goldinger, 2011; Papesh, Goldinger, & Hout, 2012). In such studies, the pupils begin to dilate within 200–300 ms following the onset of a stimulus for cognitive processing. Constriction then follows the offset of cognitive demand, on a time course similar to that of dilation. For example, in a short-term memory task in which participants held digits in working memory prior to a “recall” signal, pupils reached maximum dilation during a prereponse pause, and then gradually constricted with each recalled digit (Kahneman & Beatty, 1966).

In the present study, participants' pupil diameters increased sharply during the response preparation phase (the period directly following the response tone); this increase was consistently higher for LF than for HF words. Because this occurred in both standard and catch trials, it suggests that, regardless of the speech–motor plan, LF words demanded greater cognitive resources during processing. In fact, the frequency effect was numerically stronger during catch trials, wherein participants issued the same speech responses, regardless of the perceived word. Although the motor program for catch trials was repeated often throughout the experiment, thus producing smaller peak differences from baseline, such trials yielded large pupillary frequency

effects: Planning to speak LF words, relative to HF words, demanded greater cognitive resources. (Notably, the numerically larger effect for catch trials may have been an artifact of having fewer observations, by design.) Furthermore, even after the speech–motor plan had been executed, LF words continued to exert an influence over the availability of cognitive resources. As can be seen in Figs. 3 and 4, we observed persistent frequency effects in the postresponse phase of each trial. This phase lasted at least 4,000 ms, which is sufficient time to allow the pupils to constrict following a naming response (Kahneman & Beatty, 1966). We thus observed a lasting effect of word frequency, continuing even as participants prepared for the subsequent trial. Indeed, as we noted in note 3, pilot testing for this experiment showed that the word frequency of trial  $n$  significantly influenced the starting diameter during trial  $n + 1$ , forcing us to use a trial-by-trial baseline correction procedure on our pupil data. (Note that, in the righthand panel of Fig. 4, the uncorrected functions show a frequency-based diameter difference prior to word onset.) LF words demand more cognitive resources during processing, leaving a diminished “pool” of available resources after they are processed. Although this hypothesis requires further testing, it suggests that higher-level cognitive operations, such as reasoning or text comprehension, may suffer persistent lags when a reader encounters LF words with minimal supporting context.

In summary, the present results suggest that word frequency affects postlexical access processes, such that LF words demand greater attention in speech planning and beyond. By monitoring pupil diameters as participants perceived and prepared to speak words of varying frequencies, we observed that the cognitive demand imposed by LF words reliably exceeded that of HF words. This difference in demand occurred over an extended time course, including the postspeech “recovery” time. Current theories of frequency effects in word perception typically focus on discrete stages, from perception through behavioral output. Analogous to “spill-over” effects in reading (Rayner & Duffy, 1986; Reichle, Pollatsek, & Rayner, 2006), our results suggest that frequency may continue to affect cognitive processing after perception, lexical access, and naming are complete. The results also suggest that pupillometry offers a reliable method to estimate the cognitive demand imposed by different variables in word recognition.

**Author note** We thank Michael Hout, Kathleen Powers, Justin Wilson, and Anthony Barnhart for their support on this project. We also thank David Plaut, Keith Rayner, Michael Cortese, and Derek Besner for their valuable comments on previous versions. Support was provided by NIH/NIDCD Grant R01-DC04535-11 to the second author.

**Appendix A**

**Table 3** Matched stimulus items (high frequency/low frequency)

admit/timid	grow/crow	guy/gym	team/teem
apple/addle	refer/defer	guide/guile	taste/baste
act/cot	thread/dearth	grab/grub	coffee/toffee
accept/accent	dear/pear	hear/heal	truck/trunk
abuse/amuse	dead/deaf	park/harp	though/trough
area/aria	disk/dish	high/sigh	wage/wade
attack/attach	fire/dire	board/hoard	vote/tote
ask/asp	hill/dill	land/lend	while/whine
art/rat	hole/dole	lack/tack	write/wrist
beer/beep	pocket/docket	level/lever	would/wound
beach/beechn	drawn/drown	motion/lotion	four/fore
bear/beau	done/dose	load/loaf	lie/lye
beat/beet	hope/dope	male/lame	stock/stick
heard/beard	luck/duck	nature/mature	steel/steed
been/bean	else/seal	market/marker	state/slate
head/bead	enjoy/envoy	model/modal	strike/strife
bet/bat	end/den	milk/mild	stream/streak
big/pig	exact/enact	most/moot	strain/strait
hide/tide	eye/eve	mouse/moose	summer/simmer
bid/bib	fault/vault	must/mute	sweet/sweep
fellow/bellow	father/lather	pain/pane	sure/ruse
blood/bloom	fast/fist	past/pest	seven/sever
block/flock	lake/fake	plain/plaid	brush/shrub
blind/blink	faith/thief	piece/niece	speak/sneak
black/slack	five/hive	plant/plank	smile/slime
bitter/bitten	field/fiend	reach/cheer	sound/mound
breed/bleed	middle/riddle	rain/raid	soul/soup
foot/loot	few/pew	read/reed	still/spill
book/boom	fool/foul	ring/rink	course/coarse
bug/tug	floor/flood	rough/rouge	come/mace
bright/fright	gate/gait	rose/sore	cold/colt
gear/tear	front/frost	scale/scald	color/colon
buy/bye	glass/gloss	say/soy	cool/coop
career/careen	great/grate	saw/sew	cross/crass
care/cane	green/greed	sand/wand	county/bounty
card/cord	grant/grunt	search/starch	cry/rye
chair/roach	grand/grind	seem/seam	
save/cave	back/hack	score/scorn	

**Appendix B**

**Table 4** Analyses of variance for the baseline-corrected peak pupil data

Source	ANOVA			
	Delay (D)	<i>F</i> (3, 38)	<i>p</i>	$\eta_p^2$
Wait		62.26	<.01**	.83
Tone		8.36	<.01**	.40
Response prep		12.81	<.01**	.49
Response		9.16	<.01**	.42
Postresponse		5.45	.01*	.30
Word Frequency (WF)		<i>F</i> (1, 40)	<i>p</i>	$\eta_p^2$
Wait		2.99	.06	.08
Tone		14.66	<.01**	.27
Response prep		15.44	<.01**	.28
Response		83.46	<.01**	.68
Postresponse		15.16	<.01**	.28
Trial Type (TT)		<i>F</i> (1, 40)	<i>p</i>	$\eta_p^2$
Wait		43.17	<.01**	.52
Tone		1.05	.31	.03
Response prep		0.16	.89	.00
Response		63.30	<.01**	.61
Postresponse		2.13	.15	.05
D × WF		<i>F</i> (3, 38)	<i>p</i>	$\eta_p^2$
Wait		0.36	.78	.03
Tone		13.61	<.01**	.53
Response prep		1.49	.23	.11
Response		4.74	.01*	.27
Postresponse		1.13	.35	.08
D × TT		<i>F</i> (3, 38)	<i>p</i>	$\eta_p^2$
Wait		2.36	.08	.16
Tone		15.16	<.01**	.55
Response prep		40.29	<.01**	.77
Response		20.18	<.01**	.61
Postresponse		4.36	.01*	.26
WF × TT		<i>F</i> (1, 40)	<i>p</i>	$\eta_p^2$
Wait		3.15	.08	.07
Tone		14.47	<.01**	.27
Response prep		3.51	.06	.08
Response		31.28	<.01**	.44
Postresponse		0.51	.48	.01
D × WF × TT		<i>F</i> (3, 38)	<i>p</i>	$\eta_p^2$
Wait		0.98	.41	.07
Tone		7.29	<.01**	.37
Response prep		0.69	.57	.05
Response		5.36	.01*	.30
Postresponse		0.31	.82	.02

\* *p* < .05. \*\* *p* < .01

## References

- Andrews, S. (1989). Frequency and neighborhood effects on lexical access: Activation or search? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *15*, 802–814. doi:10.1037/0278-7393.15.5.802
- Andrews, S. (1992). Frequency and neighborhood effects on lexical access: Lexical similarity or orthographic redundancy? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *2*, 234–254. doi:10.1037/0278-7393.18.2.234
- Balota, D. A., & Abrams, R. A. (1995). Mental chronometry: Beyond onset latencies in the lexical decision task. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*, 1289–1302. doi:10.1037/0278-7393.21.5.1289
- Balota, D. A., & Chumbley, J. I. (1984). Are lexical decisions a good measure of lexical access? The role of word frequency in the neglected decision stage. *Journal of Experimental Psychology: Human Perception and Performance*, *10*, 340–357. doi:10.1037/0096-1523.10.3.340
- Balota, D. A., & Chumbley, J. I. (1985). The locus of word-frequency effects in the pronunciation task: Lexical access and/or production? *Journal of Memory and Language*, *24*, 89–106. doi:10.1016/0749-596X(85)90017-8
- Balota, D. A., Yap, M. J., Cortese, M. J., Hutchison, K. A., Kessler, B., Loftis, B., . . . Treiman, R. (2007). The English Lexicon Project. *Behavior Research Methods*, *39*, 445–459. doi:10.3758/BF03193014
- Beatty, J. (1982). Task-evoked pupillary responses, processing load, and the structure of processing resources. *Psychological Bulletin*, *91*, 276–292. doi:10.1037/0033-2909.91.2.276
- Becker, C., & Killion, T. (1977). Interaction of visual and cognitive effects in word recognition. *Journal of Experimental Psychology: Human Perception and Performance*, *3*, 389–401. doi:10.1037/0096-1523.3.3.389
- Bradshaw, J. L. (1968). Load and pupillary changes in continuous processing tasks. *British Journal of Psychology*, *59*, 265–271. doi:10.1111/j.2044-8295.1968.tb01139.x
- Brysbaert, M., & New, B. (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, Instruments, & Computers*, *41*, 977–990. doi:10.3758/BRM.41.4.977
- Chee, M. W. L., Westphal, C., Goh, J., Graham, S., & Song, A. W. (2003). Word frequency and subsequent memory effects studied using event-related fMRI. *NeuroImage*, *20*, 1042–1051. doi:10.1016/S1053-8119(03)00335-5
- Connine, C. M., Mullennix, J., Shernoff, E., & Yelen, J. (1990). Word familiarity and frequency in visual and auditory word recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *16*, 1084–1096. doi:10.1037/0278-7393.16.6.1084
- Dambacher, M., Kliegle, R., Hofmann, M., & Jacobs, A. J. (2006). Frequency and predictability effects on event-related potentials during reading. *Brain Research*, *1084*, 89–103. doi:10.1016/j.brainres.2006.02.010
- Forster, K. I., & Chambers, S. M. (1973). Lexical access and naming time. *Journal of Verbal Learning and Verbal Behavior*, *12*, 627–635. doi:10.1016/S0022-5371(73)80042-8
- Glanzer, M., & Adams, J. K. (1985). The mirror effect in recognition memory. *Memory & Cognition*, *13*, 8–20. doi:10.3758/BF03198438
- Goldiamond, I., & Hawkins, W. F. (1958). Vexiersuch: The log relationship between word-frequency and recognition obtained in the absence of stimulus words. *Journal of Experimental Psychology*, *56*, 457–463. doi:10.1037/h0043051
- Goldinger, S. D., Azuma, T., Abramson, M., & Jain, P. (1997). Open wide and say “blah!” Attentional dynamics of delayed naming. *Journal of Memory and Language*, *37*, 190–216. doi:10.1006/jmla.1997.2518
- Goldinger, S. D., He, Y., & Papesch, M. H. (2009). Deficits in cross-race face learning: Insights from eye-movements and pupillometry. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *35*, 1105–1122. doi:10.1037/a0016548
- Grainger, J., & Segui, J. (1990). Neighborhood frequency effects in visual word recognition: A comparison of lexical decision and masked identification latencies. *Perception & Psychophysics*, *47*, 191–198. doi:10.3758/BF03205983
- Granhölm, E., Asarnow, R. F., Sarkin, A. J., & Dykes, K. L. (1996). Pupillary responses index cognitive resource limitations. *Psychophysiology*, *33*, 457–461. doi:10.1111/j.1469-8986.1996.tb01071.x
- Granhölm, E., & Verney, S. P. (2004). Pupillary responses and attentional allocation on the visual backward masking task in schizophrenia. *International Journal of Psychophysiology*, *52*, 37–52. doi:10.1016/j.ijpsycho.2003.12.004
- Hauk, O., & Pulvermüller, F. (2004). Effects of word length and frequency on the human event-related potential. *Clinical Neurophysiology*, *115*, 1090–1103. doi:10.1016/j.clinph.2003.12.020
- Heitz, R. P., Schrock, J. C., Payne, T. W., & Engle, R. W. (2008). Effects of incentive on working memory capacity: Behavioral and pupillometric data. *Psychophysiology*, *45*, 119–129. doi:10.1111/j.1469-8986.2007.00605.x
- Herdman, C. M. (1992). Attentional resource demands of visual word recognition in naming and lexical decisions. *Journal of Experimental Psychology: Human Perception and Performance*, *18*, 460–470. doi:10.1037/0096-1523.18.2.460
- Inhoff, A. W., & Rayner, K. (1986). Parafoveal word processing during eye fixations in reading: Effects of word frequency. *Perception & Psychophysics*, *40*, 431–439. doi:10.3758/BF03208203
- Kahneman, D. (1973). *Attention and effort*. Englewood Cliffs, NJ: Prentice Hall.
- Kahneman, D., & Beatty, J. (1966). Pupil diameter and load on memory. *Science*, *154*, 1583–1585. doi:10.1126/science.154.3756.1583
- Karatekin, C., Couperus, J. W., & Marcus, D. J. (2004). Attention allocation in the dual-task paradigm as measured through behavioral and psychophysiological responses. *Psychophysiology*, *41*, 175–185. doi:10.1111/j.1469-8986.2004.00147.x
- Kawamoto, A. H., Liu, Q., Mura, K., & Sanchez, A. (2008). Articulatory preparation in the delayed naming task. *Journal of Memory and Language*, *58*, 347–365. doi:10.1016/j.jml.2007.06.002
- Kello, C. T. (2004). Control over the time course of cognition in the tempo-naming task. *Journal of Experimental Psychology: Human Perception and Performance*, *30*, 942–955. doi:10.1037/0096-1523.30.5.942
- Kessler, B., Treiman, R., & Mullennix, J. (2002). Phonetic biases in voice response time measurements. *Journal of Memory and Language*, *47*, 145–171. doi:10.1006/jmla.2001.2835
- Kronbichler, M., Hutzler, F., Wimmer, H., Mair, A., Staffan, W., & Ladurner, G. (2004). The visual word form area and the frequency with which words are encountered: Evidence from a parametric fMRI study. *NeuroImage*, *21*, 946–953. doi:10.1016/j.neuroimage.2003.10.021
- Kučera, H., & Francis, W. N. (1967). *A computational analysis of present-day American English*. Providence, RI: Brown University Press.
- Kuchinke, L., Võ, M. L.-H., Hofmann, M., & Jacobs, A. M. (2007). Pupillary responses during lexical decisions vary with word frequency but not emotional valence. *International Journal of Psychophysiology*, *65*, 132–140. doi:10.1016/j.ijpsycho.2007.04.004
- Manelis, J. (1977). Frequency and meaningfulness in tachistoscopic word perception. *The American Journal of Psychology*, *99*, 269–280. doi:10.2307/1422049
- McCann, R. S., & Besner, D. (1987). Reading pseudohomophones: Implications for models of pronunciation assembly and the locus of word-frequency effects in naming. *Journal of Experimental*

- Psychology. Human Perception and Performance*, 13, 14–24. doi:10.1037/0096-1523.13.1.14
- McRae, K., Jared, D., & Seidenberg, M. S. (1990). On the roles of frequency and lexical access in word naming. *Journal of Memory and Language*, 29, 43–65. doi:10.1016/0749-596X(90)90009-O
- O'Malley, S., Reynolds, M. G., Stolz, J. A., & Besner, D. (2008). Reading aloud: Spelling–sound translation uses central attention. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 34, 422–429. doi:10.1037/0278-7393.34.2.422
- Paap, K. R., & Johansen, L. S. (1994). The case of the vanishing frequency effect: A retest of the verification model. *Journal of Experimental Psychology. Human Perception and Performance*, 20, 1129–1157. doi:10.1037/0096-1523.20.6.1129
- Paivio, A., & Simpson, H. M. (1966). The effect of word abstractness and pleasantness on pupil size during an imagery task. *Psychonomic Science*, 5, 55–56.
- Papesh, M. H., & Goldinger, S. D. (2011). Your effort is showing! Pupil dilation reveals memory heuristics. In P. Higham & J. Leboe (Eds.), *Constructions of remembering and metacognition* (pp. 215–224). Palgrave Macmillan.
- Papesh, M. H., Goldinger, S. D., & Hout, M. C. (2012). Memory strength and specificity revealed by pupillometry. *International Journal of Psychophysiology*, 83, 56–64. doi:10.1016/j.ijpsycho.2011.10.002
- Porter, G., & Troscianko, T. (2003). Pupillary response to grating stimuli [Abstract]. *Perception*, 32(ECVP Abstract Suppl.), 156a.
- Porter, G., Troscianko, T., & Gilchrist, I. D. (2007). Effort during visual search and counting: Insights from pupillometry. *Quarterly Journal of Experimental Psychology*, 60, 211–229. doi:10.1080/17470210600673818
- Psychology Software Tools, Inc. (2006). E-Prime (Version 1.2). Available at [www.pstnet.com](http://www.pstnet.com)
- Rayner, K., & Duffy, S. A. (1986). Lexical complexity and fixation times in reading: Effects of word frequency, verb complexity, and lexical ambiguity. *Memory & Cognition*, 14, 191–201. doi:10.3758/BF03197692
- Reichle, E. D., Pollatsek, A., & Rayner, K. (2006). E-Z Reader: A cognitive-control, serial-attention model of eye-movement behavior during reading. *Cognitive Systems Research*, 7, 4–22. doi:10.1016/j.cogsys.2005.07.002
- Richer, F., & Beatty, J. (1985). Pupillary dilations in movement preparation and execution. *Psychophysiology*, 22, 204–207. doi:10.1111/j.1469-8986.1985.tb01587.x
- Richer, F., & Beatty, J. (1987). Contrasting effects of response uncertainty on the task-evoked pupillary response and reaction time. *Psychophysiology*, 24, 258–262. doi:10.1111/j.1469-8986.1987.tb00291.x
- Richer, F., Silverman, C., & Beatty, J. (1983). Response selection and initiation in speeded reactions: A pupillometric analysis. *Journal of Experimental Psychology. Human Perception and Performance*, 9, 360–370. doi:10.1037/0096-1523.9.3.360
- Rubenstein, H., Garfield, L., & Millikan, J. A. (1970). Homographic entries in the internal lexicon. *Journal of Verbal Learning and Verbal Behavior*, 9, 487–494. doi:10.1016/S0022-5371(70)80091-3
- Rugg, M. D. (1990). Event-related brain potentials dissociate repetition effects of high- and low-frequency words. *Memory & Cognition*, 18, 367–379. doi:10.3758/BF03197126
- Savage, G. R., Bradley, D. C., & Forster, K. I. (1990). Word frequency and the pronunciation task: The contribution of articulatory fluency. *Language & Cognitive Processes*, 5, 203–236. doi:10.1080/01690969008402105
- Scarborough, D. L., Cortese, C., & Scarborough, H. (1977). Frequency and repetition effects in lexical memory. *Journal of Experimental Psychology. Human Perception and Performance*, 3, 1–17. doi:10.1037/0096-1523.3.1.1
- Stone, G. O., & Van Orden, G. C. (1993). Strategic processes in printed word recognition. *Journal of Experimental Psychology. Human Perception and Performance*, 19, 744–774. doi:10.1037/0096-1523.19.4.744
- Van Gerven, P. W. M., Paas, F., Van Merriënboer, J. J. G., & Schmidt, H. G. (2004). Memory load and the cognitive pupillary response in aging. *Psychophysiology*, 41, 167–174. doi:10.1111/j.1469-8986.2003.00148.x
- Waters, G. S., & Seidenberg, M. S. (1985). Spelling-sound effects in reading: Time-course and decision criteria. *Memory & Cognition*, 13, 557–572. doi:10.3758/BF03198326
- Whaley, C. P. (1978). Word-nonword classification time. *Journal of Verbal Learning and Verbal Behavior*, 17, 143–154.
- Yap, M. J., Balota, D. A., Tse, C.-S., & Besner, D. (2008). On the additive effects of stimulus quality and word frequency in lexical decision: Evidence for opposing interactive influences revealed by RT distributional analyses. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 34, 495–513. doi:10.1037/0278-7393.34.3.495