University of Calgary

PRISM Repository **https://prism.ucalgary.ca**

The Vault Open Theses and Dissertations

2014-08-25

Effects of Levels-of-processing and Test-list Context on Recognition and Pupil Dilation

Taikh, Alexander

Taikh, A. (2014). Effects of Levels-of-processing and Test-list Context on Recognition and Pupil Dilation (Master's thesis, University of Calgary, Calgary, Canada). Retrieved from https://prism.ucalgary.ca. doi:10.11575/PRISM/25826

http://hdl.handle.net/11023/1693

Downloaded from PRISM Repository, University of Calgary

UNIVERSITY OF CALGARY

Effects of Levels-of-processing and Test-list Context on Recognition and Pupil Dilation

by

Alexander Taikh

A THESIS

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES

IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE

DEGREE OF MASTER OF SCIENCE

GRADUATE PROGRAM IN PSYCHOLOGY

CALGARY, ALBERTA

AUGUST, 2014

© Alexander Taikh 2014

Abstract

Recognition can be associated with experiences of recollection and/or familiarity. Bodner and Lindsay (2003) found that critical items were more likely to be experienced as recollected (vs. familiar) when tested with a set of less (vs. more) memorable items. But it remains unclear whether test context influences participants' functional definitions of recollection/familiarity or the strength of their recognition experiences. There is also a debate regarding whether pupil dilation at test reflects cognitive effort, memory strength, and/or the recreation of encoding effort. To help clarify these issues, my thesis examined the effects of levels-of-processing (LOP) and test context on recognition experiences and pupil dilation. Pupil dilation at study and test, and for recollection versus familiarity, provided partial support for all three bases of pupil dilation. In contrast, pupil dilation was not influenced by test context, suggesting context influences a judgment stage (not reflected in the pupil), rather than the recognition experience.

Acknowledgements

I thank my supervisor Dr. Glen Bodner, my supervisory committee Drs. Christopher Sears and Penny Pexman, and my external examiner, Dr. Stephen Winters. I would also like to thank my family, friends, and girlfriend for their support.

Table of Contents

Abstract ii
Acknowledgements iii
Table of Contents iv
List of Tables v
List of Figures and Illustrations vi

CHAPTER 1: INTRODUCTION 1
Test Context Effects: Evidence of an Attributional Basis for Recollect/Familiar Judgments 3
Alternative measures of recollection and familiarity 5
Pupillometry 7
The Current Study 9
Isolating the effects of LOP and test-list context on recollect/familiar judgments. 10
Using LOP and test context to clarify the basis of pupil dilation effects. 11
Using pupil dilation effects to clarify the basis of test context effects on recognition judgments. 14

CHAPTER 2: METHOD 17
Participants 17
Materials and Design 17
Procedure 17
Pupil Measurement 19

CHAPTER 3: RESULTS 20
Measuring Pupil Dilation 20
Effects of LOP on the Pupil at Study 21
Effects of LOP on Recollect/Familiar Judgments at Test 24
Effects of Old/New Item Status on the Pupil at Test 26
Effects of LOP Item Status on the Pupil at Test 26
Effects of Recollect/Familiar Judgments on the Pupil at Test 29
Effects of Test Context on Recollect/Familiar Judgments at Test 33
Effects of Test Context on Pupil Dilation for Medium LOP Items at Test 34
Does Pupil Dilation at Study Predict Recognition at Test? 37

CHAPTER 4: DISCUSSION 40
Isolating the effects of LOP and test-list context on recollect/familiar judgments 40
Using LOP and test context to clarify the basis of pupil dilation effects 41
Using pupil dilation effects to clarify the basis of test context effects on recognition judgments 44
Implications for dual-process and single-process models of memory 45
Limitations and future directions 46
Conclusion 48

REFERENCES 49

List of Tables

Table 1. Effects of Level-of-Processing (LOP) on Pupil Dilation Ratio at Study: 3 (LOP: Shallow vs. Medium vs. Deep) x 2 (Test Context: Shallow vs. Deep) x 2 (Time Period: Word vs. Response) ANOVA 23

Table 2. Effects of Old/New Item Status on Pupil Dilation Ratio at Test: 2 (Item Status: Old vs. New) x 2 (Time Period: Word vs. Response) ANOVAs 28

Table 3. Effects of Level-of-processing (LOP) Item Status on Pupil Dilation Ratio at Test: 2 (LOP: Shallower vs. Deeper) x 2 (Time Period: Word vs. Response) ANOVAs 29

Table 4. Effects of Recollect/Familiar Judgment on Pupil Dilation Ratio at Test: 2 (Judgment: Recollect vs. Familiar) x 2 (Level-of-processing: Shallower vs. Deeper) x 2 (Time Period: Word vs. Response) ANOVAs 30

Table 5. Effects of Test Context on Pupil Dilation Ratio for Medium Items at Test: 2 (Test Context: Shallow vs. Deep) x 2 (Time Period: Word vs. Response) ANOVA 35

Table 6. Effects of Test Context and Recollect/Familiar Judgment on Pupil Dilation Ratio at Test for Medium Level-of-Processing Items: 2 (Test Context: Shallow vs. Deep) x 2 (Judgment: Recollect vs. Familiar) x 2 (Time Period: Word vs. Response) ANOVA 36

Table 7. Pupil Dilation Ratio at Study as a Function of Judgment at Test: 3 (Judgment at Test: Recollect vs. Familiar vs. New) x 2 (Time Period at Study: Word vs. Response) ANOVAs 39

List of Figures and Illustrations

Figure 1. (A) Expected PDR pattern at study as a function of LOP for shallow (S), medium (M), and deep (D) LOP words, (B-D) potential PDR patterns at test as a function of LOP (E-H), potential PDR patterns at test for medium LOP words as a function of test context. 12

Figure 2. Medium words in the shallow and deep contexts: Expected PDR pattern for recollect judgments, and three potential PDR patterns for familiar judgments depending on whether the PDR pattern matches the familiarity pattern found using raw familiar judgments F(Raw), the independence remember/know correction F(IRK), or the independent recollect/familiar ratings task F(IR). 15

Figure 3. Mean pupil dilation ratios at study for shallow, medium, and deep items for each test context group word and response periods. Error bars indicate standard error. 22

Figure 4. Mean proportions of recollect and familiar judgments for item types at test in each test context. Error bars indicate standard error. 25

Figure 5. Mean pupil dilation ratios at test based on a) item status, b) old judgments, c) recollect judgments, d) familiar judgments, and e) new judgments for item types in each test context as a function of time period. Error bars indicate standard error. 27

Figure 6. Mean pupil dilation ratios at test for studied items as a function of judgment type and time period. Error bars indicate standard error. 31

Figure 7. Mean pupil dilation ratios at study as a function of judgment at test and time period. Error bars indicate standard error. 38

CHAPTER 1: INTRODUCTION

Recognition may be accompanied by recollection of specific episodic details from encoding, or by a strong feeling of familiarity without the recollection of such details (Tulving, 1985; Yonelinas, 2002). Tulving (1985) developed the remember/know judgment task to track these subjectively different recognition experiences, and to test whether they signal the operation of different underlying processes and/or memory systems. During a recognition test, participants judge each item they recognize as remember (or, as labeled here, *recollect*) or as know (or, as labeled here, *familiar*).

Since the inception of this task, multiple encoding and retrieval manipulations have been shown to have dissociative effects on recollection versus familiarity, whereas many others have parallel effects (for a review see Gardiner & Richardson-Klavehn, 2000). For example, Gardiner (1988) manipulated levels-of-processing (LOP) at study (see Yonelinas, 2002 for a review). Words encoded using a deep LOP task (focused on the meaning of items) yielded more reports of recollection, but not of familiarity, than words encoded using a shallow LOP task (focused on superficial aspects of items). A challenge for accounts of recognition memory has been to explain the extant pattern of dissociations on recollection and familiarity experiences.

According to dual-process models of recognition memory (e.g., Jacoby, 1991; Jacoby & Dallas, 1981; Yonelinas, 2002), recollection and familiarity reflect two independent retrieval processes. A consciously controlled recollection process attempts to retrieve episodic information about prior study of an item from memory. Recollection either succeeds (yielding an experience of recollection) or not—thus, recollection is a threshold process. An automatic familiarity process can also indicate prior study. Assuming that studied items generally elicit stronger feelings of familiarity than new items, participants can set a criterion along a memory

strength dimension and items exceeding this criterion are classified as familiar. Thus, deeper LOP at encoding facilitated the recollection process but had no influence on the familiarity process. Additionally, that recollection and familiarity are supported by different regions in the brain (e.g., Tulving & Markowitsch, 1998; Yonelinas, Kroll, Dobbins, Lazzara, & Knight, 1998), suggests a neural basis for the observed dissociative effects on recognition judgments.

Single-process signal-detection models posit that differences in memory strength separate judgments of recollection versus familiarity (e.g., Donaldson, 1996; Dunn, 2004; Wixted & Stretch, 2004). Items judged as recollect exceed a recollect/familiar criterion, whereas items judged as familiar do not. Dissociations are explained by changes to the memory strength distributions at study and the placement of participants' decision criteria along the strength continuum at test.

A third class of accounts, called attributional accounts, posit that the experience of recognition, and the recollect/familiar judgment applied, reflect heuristic decisions influenced by the experimental task and context (e.g., Bodner & Lindsay, 2003; Gruppuso, Lindsay, & Kelley, 1997; McCabe & Balota, 2007). Gruppuso et al. proposed the functional account of recollect/familiar judgments, which the present study also evaluated. According to the functional account, whether the memories evoked by a test item lead to the experience of recollection or familiarity depends on what they allow the participant to do in the experimental context.

In Gruppuso et al.'s (1997) paradigm, participants encoded two lists of words. At test, participants had to discriminate between List 1 and List 2 words by responding *yes* (i.e., including) to all List 2 words, and *no* (i.e., excluding) to all List 1 words and new words. All List 2 words were encoded using one type of judgment. In contrast, some List 1 words were encoded using that same judgment (making discrimination from List 2 words *difficult*) and others were

encoded using a different judgment (making discrimination *easy*). Correctly excluded List 1 words contributed to estimates of *recollection*. Incorrectly included List 1 words contributed to the estimate of *familiarity* because participants recognized the words but failed to recollect that they were studied on List 1. Recollection was greater when discrimination was easy (vs. difficult), whereas familiarity was greater when discrimination was difficult (vs. easy). Thus, retrieving information about the encoding task of List 1 items was more likely to contribute to recollection when that information was useful in figuring out the source list. By the functional account, then, the same retrieved information can result either in a judgment of recollection or familiarity depending on the task.

Test Context Effects: Evidence of an Attributional Basis for Recollect/Familiar Judgments

Consistent with the notion that a retrieved memory can result either in an experience of recollection or familiarity, Bodner and Lindsay (2003) found that the test context influenced whether items from a given encoding condition were judged as recollect or as familiar. I used this paradigm in my thesis. In their Experiment 1, participants in the shallow context studied one list of words using a medium level-of-processing (LOP) task (Is this something people commonly use?), and studied another list of words using a shallow LOP task (Does the word contain the letter "A"?). Participants in the deep context studied one list using the medium LOP task, but the other list using a deep LOP task (Is this something you would want if you were stranded on a desert island?). At test, both context groups were presented with the words they studied along with new ones, and were asked to judge each one as recollect, familiar, or new. Medium words were more likely to be judged as recollect and less likely to be judged as familiar in the shallow (vs. deep) context.

According to the functional account, test context changed how participants defined recollection and familiarity. To more directly test this, participants in Experiment 4 were asked to provide the strongest memory accompanying each recollect judgment. Experiment 4 replicated the context effect. Critically, participants given the shallow (vs. deep) context were more likely to emphasize recollecting spontaneous thoughts and associations experienced at study. In contrast, participants in the deep (vs. shallow) context were more likely to emphasize recollecting the encoding task experienced at study.

However, both the pair of LOP conditions was different at both study and test in Bodner and Lindsay (2003, Experiment 1 and 4). Thus, it is possible that different study experiences rather than different test contexts, were the basis of these effects. Indeed, McCabe and Balota, (2007) provided evidence that study context can also influence recognition judgments. Fortunately, Bodner and Lindsay (2003, Experiment 2) provided evidence for a role of test context. Experiment 2 was the same as the other experiments except only medium LOP and new words were presented at test. Thus, the study context varied but the test context was held constant. Here medium items did not differ in their rates of recollect and familiar judgments across the different study contexts.

Finally, Experiment 3 of Bodner and Lindsay (2003) examined whether test list context modulated participants ability to recollect specific types of information, in this case list-source information, for the medium items. Here, participants made source judgments (i.e., List 1 or List 2 or new?), instead of recognition judgments. Source accuracy for medium items was the same across test contexts, suggesting that test context affects what information participants *use* rather than what information they *can access*. However, although context did not affect access to source information, it remains possible that context might affect the overall amount of

information that comes to mind. Consistent with this possibility, Bodner and Lindsay (2003) found that overall recognition (sum of remember and know judgments) for medium items was greater in the shallow (vs. deep) test context across experiments. Thus Bodner and Lindsay could not rule out the possibility that context also affects total recollection or overall memory strength.

Alternative measures of recollection and familiarity

A criticism of the binary judgment task is that recognition experiences based on recollection are treated as mutually exclusive to ones based on familiarity. If participants experience *both* recollection and familiarity they will likely choose to report a recollect experience given the task instructions define familiarity as the absence of recollection. Manipulations that increase the proportion of recollect judgments for a set of items will thus work to limit the proportion of familiar judgments, even if the experience of familiarity was not affected. This limitation might explain the counterintuitive finding that a deeper LOP sometimes yields less familiarity than a shallow LOP task (for a review, see Yonelinas, 2002).

According to dual-process theory, the binary judgment task will tend to underestimate familiarity. The independence remember/know (IRK) estimate of familiarity is computed to attempt to correct for this underestimation (Yonelinas & Jacoby, 1995). The numerator of the IRK estimate is the proportion of familiar judgments and the denominator is the proportion of trials in which a recollect judgment was not made, thus $IRK = F/(1-R)$. When Bodner and Lindsay (2003) applied the IRK correction, test context did not influence familiarity (cf. “raw” know judgments). Thus, medium items were more likely to be recollected in a shallow context but were not less likely to be experienced as familiar. As discussed in Tousignant, Bodner, and Arnold (2014), although the correction may help reduce underestimation of familiarity, it does not address the underlying mismatch with the structure of the binary judgment task, namely that

recollect and familiar experiences are treated as *mutually exclusive* even though dual-process theory treats the underlying processes as *independent*.

Higham and Vokey (2004) developed a novel task to eliminate the mutual exclusivity of recollect and familiar responses. In their independent-ratings task, participants separately rate the extent to which they experience recollection and familiarity for a given test item. Using independent ratings, participants are free to indicate having both, one, or neither of the two recognition experiences, and to different degrees, on each trial. To date, studies using the independent rating task have typically yielded parallel effects on recollect and familiar ratings, using the same variables that yield dissociations with binary judgments (e.g., Brown & Bodner, 2011; Higham & Vokey, 2004; Kurilla & Westerman, 2008; Tousignant & Bodner, 2012; Tousignant et al., 2014).

Of relevance to the present study, using independent ratings Tousignant and Bodner (2012) found a parallel effect of test context rather than a dissociation: Medium items received higher ratings of both recollection and familiarity in the shallow (vs. deep) context. They argued that their results are compatible with the functional account assuming the definitions of remembering and knowing participants formed in the shallow (vs. deep) context made medium and new items feel both more recollected and more familiar. The effect of test context on recollect/familiar judgments thus depends on the task (judgments vs. ratings). In sum, the binary judgment task suggests that the shallow (vs. deep) test context increases recollection and decreases familiarity, the IRK estimate suggests that it increases recollection but has no influence on familiarity, and the independent ratings task suggests it increases both experiences.

Pupillometry

Given the discrepancies across binary judgment and independent rating tasks outlined above, a central goal of my thesis was to see whether the pupil can help triangulate how test context affects recognition experiences. The pupil responds to external sensory events, for example, constricting in response to luminance or dilating in response to auditory stimuli (Beatty & Lucero-Wagoner, 2000). Additionally, the pupil responds to internal processes such as emotional arousal (Hess and Polt, 1960), although see Hess and Polt (1964; Vo et al., 2008). Non-reflexive pupil dilation has been used as an indicator of the amount of cognitive processing evoked by a task, independently of reflexive pupil dilation (Beatty & Lucero-Wagoner, 2000). Previously, greater pupil dilation has been found to reflect greater processing load induced, for example, by mental arithmetic (Hess & Polt, 1964), demand on short term memory (Clark & Johnson, 1970; Kahneman & Beatty, 1966), or effort required to transform an ambiguous sentence to an unambiguous one (Schluroff et al., 1986). Additionally, Ariel and Castel (2014) found that pupil dilation also reflects the allocation of attentional resources (see also Kahneman & Peavler, 1969), suggesting dilation can be evoked by conscious decision-making strategies.

Characteristics of the pupil's response support its utility as a measure in cognitive tasks. For example, the pupil's response has been shown to be independent of its baseline diameter (Bradshaw, 1969; Granholm, Asarnow, Sarkin, & Dykes, 1996), and it is not influenced by the previous trial (Naber, Frassle, Rutishauser, & Einhauser, 2013). Other characteristics point to the need to exercise care when using pupil dilation measures. For example, the onset of the pupil response is task dependent. Principal component analyses have revealed onsets ranging from 300 ms to over 2000 ms depending on the task (Jainta & Baccino, 2010; Kuchinke, Vo, Hofmann, & Jacobs, 2007; Nuthman & van der Meer, 2005).

The pupil may provide a useful measure of long-term memory, as dilation may reflect the effort associated with retrieval, organization, and production of responses in a recognition task (Beatty & Kahneman, 1966). Interestingly, studied (vs. non-studied) nonsense words have been found to elicit greater pupil dilation at test (Gardner, Mo, & Borrego, 1974; Gardner, Mo, & Krinsky, 1974). However, as discussed below, the factor(s) that influence pupil dilation during recognition are not yet fully agreed upon. More recently, Vo et al. (2008) found that the pupil tends to dilate more when participants correctly recognize a studied item versus correctly reject new items at test (see also Heaver & Hutton, 2011; Montefinese, Ambrosini, Fairfield, & Mammarella, 2013; Naber et al., 2013; Otero, Weekes, & Hutton, 2011). Vo et al. termed this finding the *pupil old/new effect*. This effect is consistent with ERP studies showing greater positive waveforms for studied than new items at test (Rugg et al., 1995; Wiese & Daum, 2006). Vo et al. (2008) attributed the pupil old/new effect to greater cognitive effort being required for classifying studied words as old using "consciously accessible information about prior occurrence of the test item" (p. 138) (i.e., *recognition effort*), than for classifying non-studied items as new.

In contrast, Otero et al. (2011) claimed that the shows greater dilation in response to greater *memory strength* evoked by a test item (cf., Rijn, Dalenberg, Borst, & Sprenger, 2012) instead of the cognitive effort used to recollect it (see also Montefinese et al., 2013). In their Experiment 1, Otero et al. replicated the pupil old/new effect both when response at test was ignored (i.e., comparing studied vs. new items), and when correct recognition was compared to correct rejection (as done by Vo et al., 2008). But critically, Otero et al. further found that dilation was greater for recollected words than familiar words (i.e., *the pupil recollect/familiar effect*). In their Experiment 2, Otero et al. (2011) tested the possibility that the pupil responds to

cognitive effort required to classify the word as old. To this end, LOP was manipulated at study. If the pupil responds to the effort required to recognize items, then dilation should have been greater for shallow than deep LOP items. Instead, consistent with a memory strength account, pupil dilation was greater for deep LOP words.

Other evidence suggests that the cognitive processing performed at encoding may come to mind again when an item is encountered again at test (*re-experiencing of encoding*). According to the Selective Construction and Preservation of Experience (SCAPE) framework (Whittlesea, 1997; Whittlesea, 2002a; Whittlesea, 2002b), recognition first involves production of information about the test item followed by an evaluation of that information, leading to an attribution (i.e., a subjective recognition experience) which then guides task responses. Papesh et al. (2012), had participants listen to words and nonwords at study before making confidence ratings on a recognition test. Pupil dilation was measured at study and at test. Two key findings support the argument that participants recreated study encoding at test. First, high confidence ratings to studied items were associated with greater pupil dilation than low confidence ratings at both study and test. Second, pupil dilation was greater for nonwords (vs. words) at both study and test. Papesh et al. took this pattern as an indication that during recognition participants re-create the processes they applied during encoding.

The Current Study

My thesis had three main goals. The first goal was to isolate the behavioral effects of LOP and test context on recollection and familiarity judgments using a confound-free design (cf. Bodner & Lindsay, 2003; Tousignant & Bodner, 2012; Tousignant et al., 2014). Using this improved design, the second goal was to clarify what the pupil responds to at study and at test. The third goal was to use pupil measurements to help understand the basis of recollection and

familiarity judgments, and in particular how these judgments are influenced by test context. Each of these aims is next described in more detail.

Isolating the effects of LOP and test-list context on recollect/familiar judgments.

The present study better isolated the effects of both LOP and test context reported in Bodner and Lindsay (2003) by eliminating two confounds present in their design. In their study, shallow words displayed for 500 ms, medium words for 1000 ms, and deep words for 2000 ms (a study duration confound), and moreover shallow words were read silently whereas medium and deep words were read aloud (a production effect confound, given the memory advantage for words said aloud versus read silently; e.g., MacLeod, Gopie, Hourihan, Neary, & Ozubko, 2010). These confounds were intentional, to ensure three different levels of recognition (i.e., shallow < medium < deep). However, it was critical to control these confounds in my thesis to isolate the effect of LOP on pupil dilation.

Second, test context was confounded with study context in Bodner and Lindsay (2003). Participants in the shallow context received medium and shallow words at study and test, while participants in the deep context received medium and deep words at study. This confound raises the possibility that differences in recollect/familiar judgments arose from different study experiences rather than different test contexts (cf. Bodner & Lindsay, 2003, Experiment 2; McCabe & Balota, 2007). This confound also leads to the possibility that study context may influence the strength with which medium items were encoded, but recollect/familiar judgments of medium items at test may be based on the evaluation of different aspects of memories.

The aforementioned production effect confound on assessing LOP effects may also have confounded the test context effects. In the shallow context, the medium words were read aloud, while shallow words were read silently. In the deep context, the medium and deep words were

both read aloud. Medium words in the shallow context may have been experienced as more distinct because they were read aloud at study while the shallow words were not. This may have led to more recollect judgments for medium words in the shallow context. In contrast, in the deep context both the medium and the deep words were produced so there was no difference in their distinctiveness with respect to production.

In my study, all words were presented for 2 s at study and all were read silently. Study context was also held constant, as all participants studied all three lists of words (shallow, medium, and deep) regardless of which test context they received. This design ensured that the effects of LOP and test context could be isolated (cf. Bodner & Lindsay, 2003; Tousignant & Bodner, 2012; but see Bodner & Richardson-Champion, 2007). Thus, I attempted to replicate the effects of LOP and test context on recollect/familiar judgments reported in Bodner and Lindsay while eliminating these confounds.

Using LOP and test context to clarify the basis of pupil dilation effects.

No study has examined the effects of LOP on the pupil at study (only at test; Otero et al., 2011). I expected pupil dilation at study would reflect the effort needed for the encoding task, based on with research suggesting that cognitive load increases pupil dilation (Beatty & Lucery-Wagoner, 2000). Specifically, I expected to find the following pattern in terms of pupil dilation: shallow < medium < deep (see Figure 1, Panel A).

The set of factors that influence pupil dilation during retrieval are not yet clear. One possibility is that pupil dilation at test reflects re-experiencing of effort used at encoding. Building on the SCAPE framework, Papesh et al. (2012) suggested that at test, participants call to mind their study experience, including the cognitive effort they exerted. If so, pupil dilation should be greater for deeper LOP items at test (see Figure 1, Panel B). Alternatively, pupil

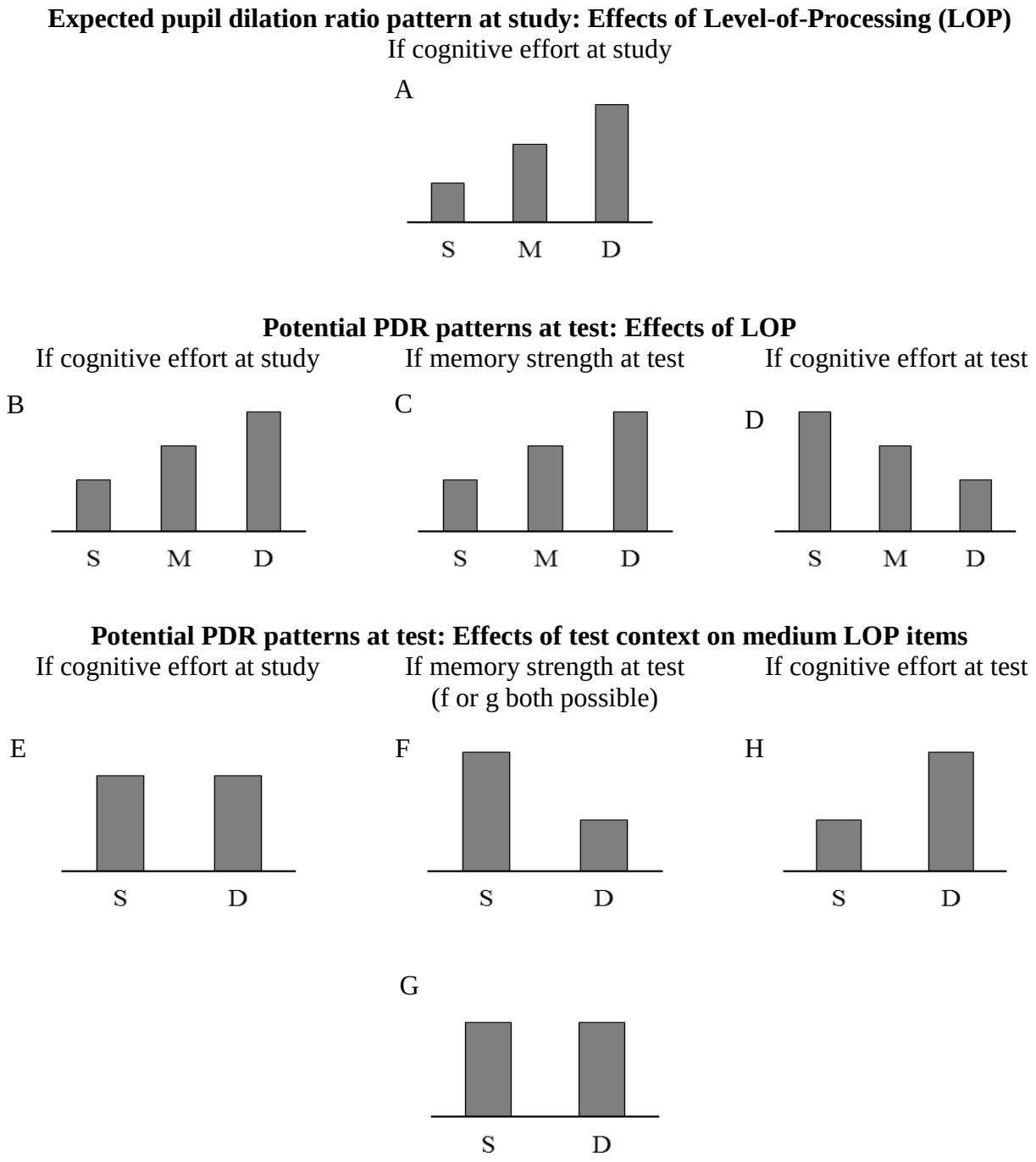

Figure 1. (A) Expected PDR pattern at study as a function of LOP for shallow (S), medium (M), and deep (D) LOP words, (B-D) potential PDR patterns at test as a function of LOP (E-H), potential PDR patterns at test for medium LOP words as a function of test context.

dilation at test may reflect the memory strength trace for that word (Otero et al., 2011). If so, once again pupil dilation should be greater for deeper LOP items at test (Figure 1, Panel C). A third possibility is that the pupil responds to the recognition effort required to recognize words at test (Vo et al., 2008). If so, deeper LOP items should be easier to recognize at test, thus yielding smaller pupil dilation (Figure 1, Panel D). However, Otero et al. (2011, Experiment 2) provided evidence against this possibility: Pupil dilation at test was greater for deeper LOP items.

In contrast, all three possibilities are consistent with the old/new pupil effect (Otero et al., 2011; Vo et al., 2008): reconstitution of encoding would lead to greater dilation when viewing old (vs. new) items, old items should elicit stronger recognition experiences than new items, and recognizing old items should require more cognitive effort and hence yield more pupil dilation than rejecting new words. Thus, while I expect to replicate the old/new effect, the effects of LOP and test context on the pupil were expected to be more informative in clarifying what pupil dilation indexes during recognition.

The two accounts for the expected LOP effects (Figure 1, Panel B vs. C) can potentially be differentiated by their predictions for the effects of test context on the pupil's response to medium items. If pupil dilation is driven by re-experiencing of encoding, then it should be insensitive to test context (Figure 1, Panel E) given the encoding context was held constant. If pupil dilation is driven by memory strength at test, then recognition of medium items might be experienced more strongly in the shallow context (see Figure 1, Panel F), consistent with the small increase in overall recognition reported by Bodner and Lindsay (2003). On the other hand, if context affects participants' functional definitions rather than the recognition experience itself (see Figure 1, Panel G), as Bodner and Lindsay emphasized, then these two accounts cannot be differentiated using the test context manipulation. Finally, if the pupil responds to the recognition

at test, then pupil dilation should be greater for medium words in the deep (vs. shallow) test context (Figure 1, Panel H) where medium words are less distinctive, and thus harder to recognize.

Using pupil dilation effects to clarify the basis of test context effects on recognition judgments.

My third goal was to use pupil dilation to help clarify how test context affects recognition judgments. These effects may be driven by a change in the definition of what constitutes recollection or by a change in the perceived strength of medium items; the functional account allows both possibilities (see Bodner & Lindsay, 2003). Additionally, the nature of the recognition measure (binary judgments vs. IRK-corrected estimates of familiarity vs. independent ratings) has been shown to influence recognition judgments. Examining pupil dilation thus provided an independent index of recognition experiences that might be useful for triangulating along these different behavioral measures and outcomes.

To this end, I compared pupil dilation for medium items across test contexts. I expected to replicate the recollect/familiar pupil effect found by Otero et al. (2011). Moreover, given that judgment and rating tasks both typically show an influence of context on recollection (cf. Tousignant et al., 2014), pupil dilation might be greater for recollected medium words in the shallow (vs. deep) test context (see *R* in Figure 2). Alternatively, the absence of a context difference in pupil response to recollected medium items would imply that context affects the likelihood of having a recollect experience but does not affect the strength of those experiences. Such a pattern would be consistent with the context effect on recollection reflecting qualitatively (vs. quantitatively) different definitions of recollection and familiarity (as per the functional account of Bodner & Lindsay, 2003).

Potential pupil dilation ratio (PDR) patterns at test:

Effects of test-list context on recognition judgments for medium words

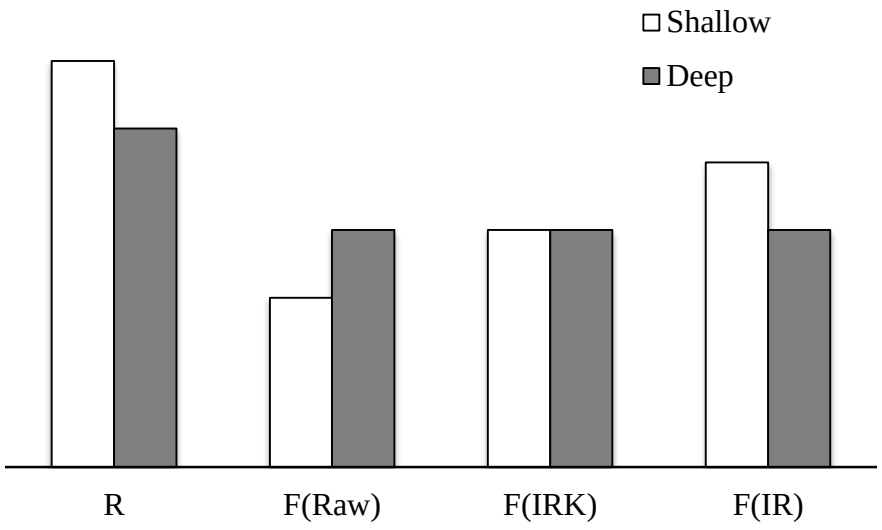

Figure 2. Medium words in the shallow and deep contexts: Expected PDR pattern for recollect judgments, and three potential PDR patterns for familiar judgments depending on whether the PDR pattern matches the familiarity pattern found using raw familiar judgments F(Raw), the independence remember/know correction F(IRK), or the independent recollect/familiar ratings task F(IR).

How test context influenced familiarity, on the other hand, has been found to differ depending on the measure used. If the raw rate of familiar judgments accurately represents the effect of test context on familiarity, then pupil dilation should be lower for medium items judged familiar in the shallow (vs. deep) test context (see *F(Raw)* in Figure 2). This would suggest that medium words lead to more recollection but less familiarity in the shallow context. On the other hand, if pupil dilation is equal for familiar medium items across test context (see *F(IRK)* in

Figure 2), then IRK-corrected estimates of familiarity would accurately describe the effect of test context on familiarity. Lastly, if independent ratings most accurately characterize the effects of test context, then the pupil response should be greater in the shallow (vs. deep) test context (see *F(IR)* in Figure 3). This would suggest that medium items are more familiar in the shallow (vs. deep) test context.

CHAPTER 2: METHOD

Participants

University of Calgary undergraduates participated for course credit and were randomly assigned to either the shallow or deep test context group (36 per group). Three participants were replaced whose proportion of recollection judgments for new items was more than 2 standard deviations above average. An additional six participants were replaced for having mean pupil dilation ratios at study more than 2 standard deviations above average.

Materials and Design

The stimuli were 12 filler and 100 critical five-letter words selected from Bodner and Lindsay (2003). Using five-letter words helped minimize luminance differences that affect pupil dilation (Goldinger & Papesh, 2012). Of the 100 critical words, for each participant 20 were randomly assigned to each of the shallow, medium, and deep study lists; the remaining 40 served as distractors at test. Two filler words were randomly assigned to the beginning and end of each study list as primacy and recency buffers. The order of the shallow, medium, and deep lists at study was counterbalanced. The order of items on each studied list was randomized. At test, a randomized mixture of 40 studied words (the medium list with either the shallow or deep list) and 40 new words was presented.

Procedure

Participants were tested individually using a computer and eye tracker, and they responded using left/right buttons on a response box. Participants first underwent an eyetracking calibration and validation procedure lasting about 5 minutes. Words and cues were presented in black 22 pt Arial font against a grey (RGB = 150) background (Papesh & Goldinger, 2012).

Participants were told they would study three word lists in turn, one word at a time, for an unspecified memory test. They were asked not to blink while the words were on the screen. They read each word silently before making a yes/no response. Prior to each list, a study instructions screen was presented. For the shallow list, participants decided whether the word contained the letter “A”. For the medium list, they decided whether the word is commonly used in English. For the deep list, they decided whether the word represents something they would like to have if they were stranded on a desert island. Participants were asked not to rehearse.

Each study trial began with a fixation signal at the center of the screen consisting of two arrows pointing inward (> <), which disappeared when participants pressed either response box button to initiate the trial. A fixation cross (+) appeared in the center of the screen for 2 s and was then replaced by a blank screen for 500 ms, followed by the study word in lowercase for 2 s. When the cue words *yes* and *no* (about 5 cm to the left and right of fixation respectively) replaced the study word, participants made their response using the left/right buttons of the response box. Study responses were not recorded.

After the study phase, participants took a break (about 1 min), then underwent a second calibration and validation procedure. The recognition judgment instructions, based on Brown and Bodner (2011, Appendix A), were presented on the screen and reviewed by the experimenter. In brief, participants were told that they would see words on the screen one at a time, some of which were old (i.e., from one of the study lists) and some of which were new (i.e., not from one of the study lists). They were told to first decide whether each word was *old* or *new*. If they responded *old* they would then be prompted to classify their recognition experience for the word as either *recollection* or *familiarity*. Participants were instructed to make a *recollect* judgment if their recognition was accompanied by recollection of specific detail from the encoding of the

word, or a *familiar* judgment if their recognition was accompanied by a strong feeling of familiarity but no conscious recollection of encoding the word. Participants described each judgment in their own words. The test phase began once the experimenter was satisfied that the instructions were understood.

Each test trial followed the same parameters used at study except the cues *old* and *new* appeared after word offset. If an *old* response was made, the cues *recollect* and *familiar* appeared on the left and right sides of screen, respectively, and participants pressed the appropriate button to classify their recognition experience.

Pupil Measurement

An EyeLink II eye tracker (*SR Research*, Kanata, ON, Canada) measured the diameter of the right pupil throughout the study and test phase. The eye tracker was equipped with a desk-mounted camera. Participants used a chinrest during the study and test phases.

CHAPTER 3: RESULTS

Measuring Pupil Dilation

The pupil measurement values provided by the software were in arbitrary units related to the number of pixels in the pupil image. Thus, a baseline pupil diameter was established for every trial as a basis from which to establish the pupillary response to the words (Beatty & Lucero-Wagoner, 2000). The baseline chosen was the peak during the first 250 ms of the word's presentation (as in Otero et al., 2011). This baseline was selected based on preliminary analysis of the pupil's response onset during study across 50 ms bins. Pupil dilation peaked during the word period (approximately 800-1300ms after the word was presented), well after the baseline.

The magnitude of pupil dilation relative to baseline was measured during two separate time periods at study and at test: During the remaining 1750 ms of the presentation of the word (*word period*) and during the response period after the word disappeared until responses were made (*response period*). The decision to include the response period (cf. Otero et al., 2011) was made after data collection because although pupil dilation peaked during the word period, it clearly continued after the presentation of the word when participants were making recognition judgments. During the word period, a pupil dilation ration (PDR) was calculated by dividing the peak dilation during last 1750 ms of the word's presentation by the baseline. During the response period, a PDR was calculated by dividing the peak dilation during the remainder of the study or test trial (i.e., after the word's presentation) by the baseline.

Peak pupil dilation was used, rather than the mean dilation, for three reasons. First, peak dilation was used in Otero et al. (2011), which provided the method I most closely followed. Second, given that pupil dilation could begin anytime during the word period, using the mean of the last 1750 ms might underestimate the pupil's response (see Jainta & Baccino, 2010;

Kuchinke, Vo, Hofmann, & Jacobs, 2007; Nuthmann & van der Meer, 2005). This was a particular concern given that the latency of the pupil's response depends on the experimental manipulations (Beatty & Lucero-Wagoner, 2000). Third, peak dilation (cf. mean dilation) is not dependent on the number of contributing data points (Beatty & Lucero-Wagoner). This was especially relevant during the response period, when participants made either one or two judgments of variable latencies. Finally, the number of contributing data points would also be affected by blinking, which also varies across trials/participants. Even so, preliminary analyses based on mean dilations closely matched those based on peak dilations.

Pupil size cannot be measured during blinks thus pupil dilation measures were based on non-blink samples. At study, blinks occurred on 24.6% of the trials during the word period, and on 12.5% of the trials during the response period. However, only 0.6% and 0.3% of these respective periods contained three or more blinks (Jainta & Baccino, 2010). At test, blinks occurred on 21.5% of trials during the word period, and on 32.3% of trials during the response period. Only 0.5% and 2.2% of these respective periods contained three or more blinks. Trials with blinks were not eliminated but 0.49% of all trials across study and test had no recorded measurements and thus did not contribute to the analyses.

Effects of LOP on the Pupil at Study

Figure 3 shows the mean PDRs at study; these means were analyzed as laid out in Table 1. A significance level of .05 was used for all analyses, and for brevity only the key effects and other significant effects are described in the text. There was a main effect of LOP at study such that deeper LOP tasks were associated with larger PDRs (shallow < medium < deep; *p*s < .05). To my knowledge, this is the first demonstration of an effect of LOP on pupil dilation at study.

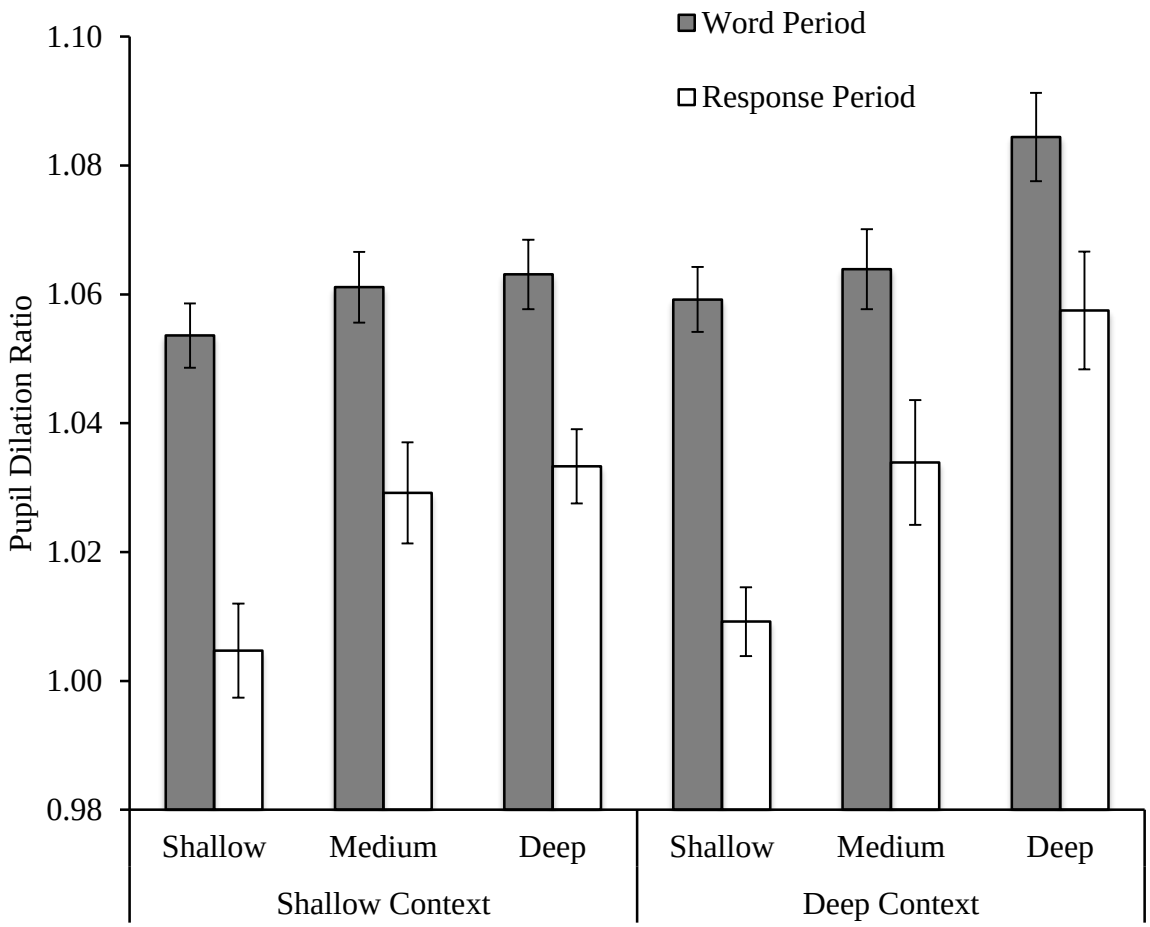

Figure 3. Mean pupil dilation ratios at study for shallow, medium, and deep items for each test context group word and response periods. Error bars indicate standard error.

Table 1

Effects of Level-of-Processing (LOP) on Pupil Dilation Ratio at Study: 3 (LOP: Shallow vs. Medium vs. Deep) x 2 (Test Context: Shallow vs. Deep) x 2 (Time Period: Word vs. Response) ANOVA

Effect (*df*)	*F*	*MSE*	η^2
LOP (2, 140)	19.14*	0.001	.22
Test Context (1, 70)	2.17	0.005	.03
Time Period (1, 70)	181.75*	0.001	.72
LOP x Test Context (2, 140)	2.77	0.001	.04
LOP x Time Period (2, 140)	14.30*	0.000	.17
Test Context x Time Period (2, 140)	0.05	0.001	.00
LOP x Test Context x Time Period (2, 140)	0.11	0.000	.00

Note: * = $p < .05$.

Although the LOP effects in Figure 1 appear more pronounced in the deep context group, the test context main effect and the interaction of LOP with test context did not reach significance.

The main effect of time period reflected greater pupil dilation during the word period than the response period. This difference could reflect changes in luminance on the screen across the two periods (when the word was replaced with response cues), or it could reflect the LOP tasks being completed within the 2 s of each word's presentation. Contrary to the latter possibility, a significant interaction between LOP and time period reflected stronger LOP effects during the response period. This interaction highlights the value of measuring pupil dilation

during the response period when responses are executed, and not just during word presentation (cf. Otero et al., 2011). Following up the interaction, one-way ANOVAs comparing shallow, medium, and deep items revealed significant LOP effects during the word period, $F(2, 142) = 11.14$, $MSE = 0.001$, $\eta^2 = .14$, and the response period, $F(2, 142) = 20.47$, $MSE = 0.001$, $\eta^2 = .22$.

In sum, at study the pupil dilated more during deeper LOP tasks, suggesting the pupil may be sensitive to cognitive effort during study, as Otero et al. (2011) found for LOP effects on the pupil at test (see also Beatty & Lucero-Wagoner, 2000). Of course, larger dilations for deeper LOP tasks could reflect a different basis, such as greater elaborative processing or episodic memory recruitment. Below, I report additional analyses testing whether pupil dilation at study was predictive of pupil dilation at test, and of the recognition judgment received at test.

Effects of LOP on Recollect/Familiar Judgments at Test

Study task order did not modulate the key behavioral findings, hence for brevity this factor was not included in the reported analyses. Figure 4 provides the mean proportion of recollect and familiar judgments at test for each item type in each test context. As is clear from Figure 4, the hit rates for studied items always exceeded the false alarm rates for new items (*p*s < .05). Overall recognition (old judgments) was near ceiling for medium and deep LOP items, hence analyses of LOP effects focused on recollect and familiar judgments (in separate ANOVAs, given the two judgment rates are not independent), and IRK estimates of familiarity.

Deeper LOP items in each test context received a higher proportion of recollect judgments (shallow < medium, medium < deep), but a lower proportion of familiar judgments (shallow > medium, medium > deep) (*ps* < .05). The mean IRK estimate of familiarity (excluding participants with recollect rates of 1 and/or familiar rates of 0) was significantly higher for medium (vs. shallow) LOP items in the shallow context (.83 vs. .52), $t(34) = 6.83$,

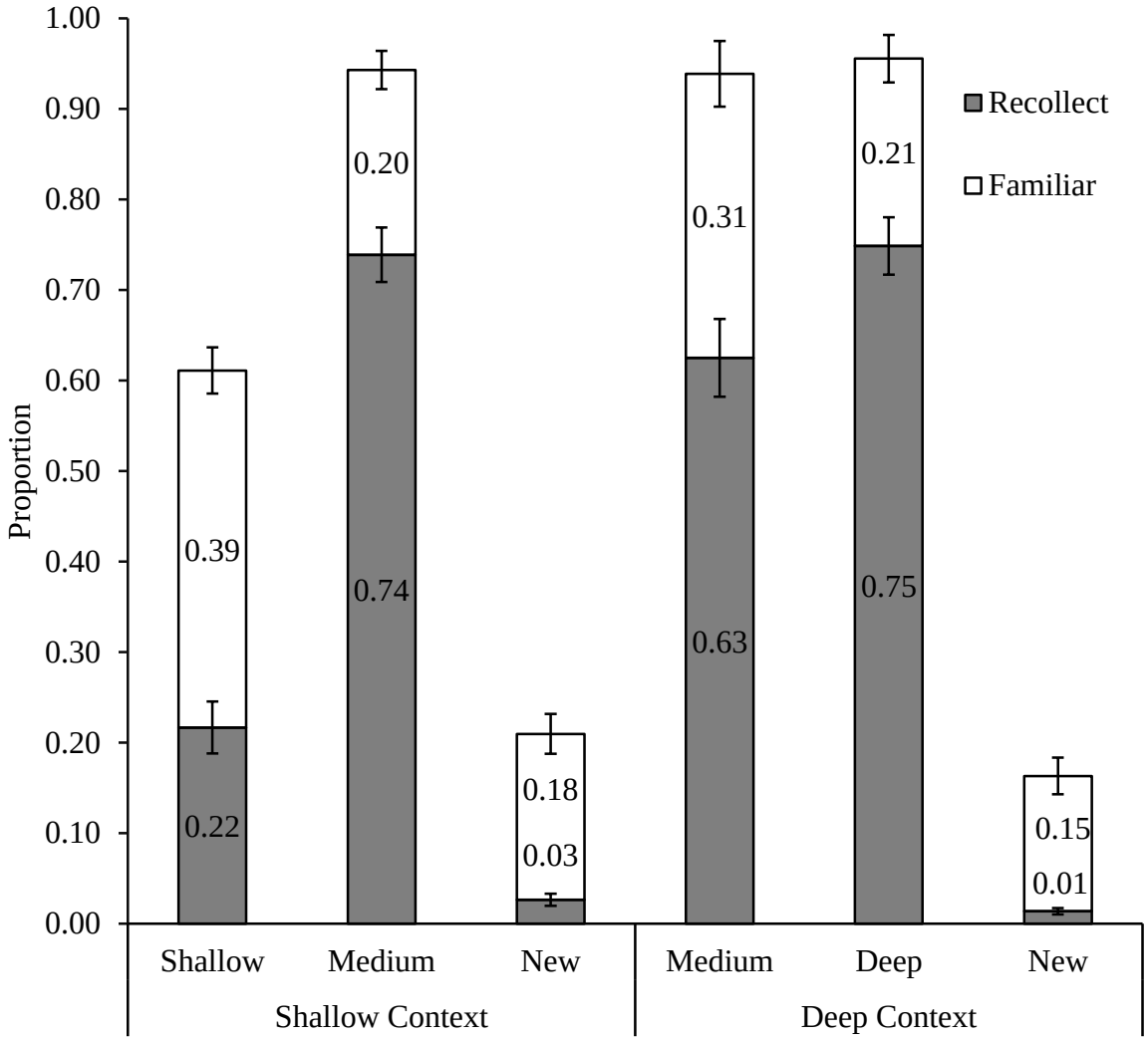

Figure 4. Mean proportions of recollect and familiar judgments for item types at test in each test context. Error bars indicate standard error.

SEM = 0.046, but not for deep (vs. medium) LOP items in the deep test context (.87 vs. .88), $t < 1$, as in Bodner and Lindsay (2003). Thus, LOP influenced recollection and familiarity in the usual manner (see Yonelinas, 2002) even when study duration and production were controlled (cf. Bodner & Lindsay, 2003; Tousignant & Bodner, 2012; Tousignant et al., 2014).

Effects of Old/New Item Status on the Pupil at Test

Figure 5 (Panel A) provides the mean PDRs at test based on item status for each time period and test context, ignoring the judgment items received. The analyses comparing the mean PDRs for each type of studied item versus new items, in each context, are provided in Table 2. In each case, the pupil dilated more for studied items than new items, replicating earlier reports of an old/new pupil effect (e.g., Otero et al., 2011; Vo et al., 2008).

There was also an interaction between item status and time period: PDR increased across time periods for studied items but decreased for new items (see Figure 5, Panel A). Pairwise t-tests comparing PDRs in the word (vs. response) periods (in each test context, and for each level of item status), confirmed these differences (ps < .05), except for shallow items, $t < 1$. This novel pattern is further considered in the discussion.

Effects of LOP Item Status on the Pupil at Test

Table 3 provides the analyses of pupil dilation at test for the deeper (vs. shallower) LOP items in each test context, again ignoring the judgment made. Pupil dilation was greater for medium (vs. shallow) LOP items in the shallow context, $t(35) = 2.19$, $SEM = 0.008$, replicating Otero et al. (2011), but was not greater for deep (vs. medium) LOP items in the deep context, $t < 1$ (see Figure 5, Panel A). However, the latter null effect could reflect ceiling recognition in the deep context. The main effects of time period in each test context reflected greater pupil dilation for studied items during the response period than the word period. The increase in PDR across time periods at test contrasts with the decreases at study. It appears that the LOP tasks are typically completed during the word period (save for the yes/no response), whereas the processing associated with recognition continues to build during the response period.

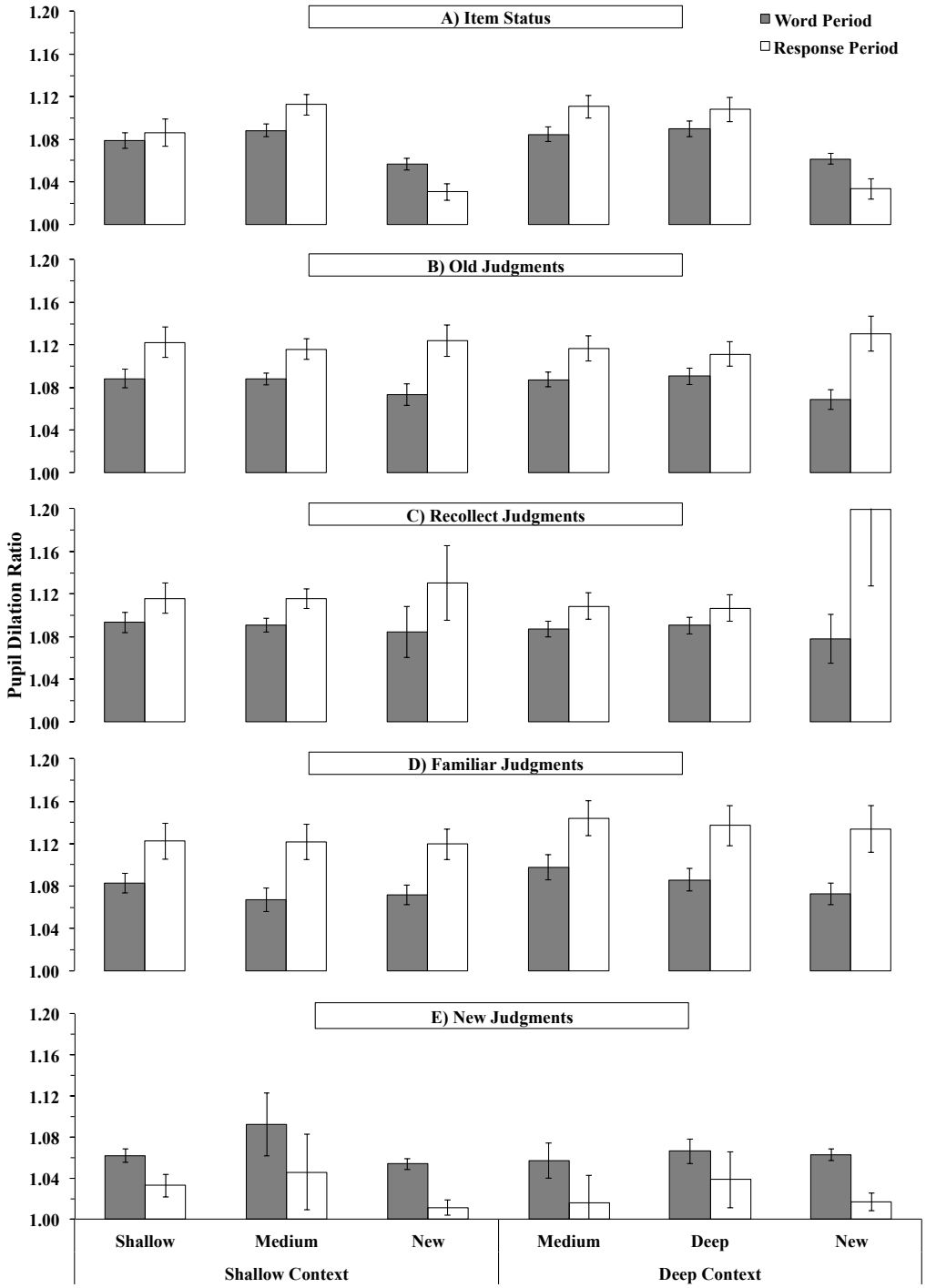

Figure 5. Mean pupil dilation ratios at test based on a) item status, b) old judgments, c) recollect judgments, d) familiar judgments, and e) new judgments for item types in each test context as a function of time period. Error bars indicate standard error.

Table 2

Effects of Old/New Item Status on Pupil Dilation Ratio at Test: 2 (Item Status: Old vs. New) x 2 (Time Period: Word vs. Response) ANOVAs

	Shallow Context			Deep Context		
Level-of-processing(LOP)/Effect (*df*)	*F*	*MSE*	η^2	*F*	*MSE*	η^2
Shallower LOP						
Item Status (1, 35)	49.76*	0.001	.59	90.57*	0.001	.72
Time Period (1, 35)	2.52	0.001	.06	0.04	0.001	.00
Item Status x Time Period (1, 35)	27.57*	0.000	.44	72.95*	0.000	.68
Deeper LOP						
Item Status (1, 35)	77.43*	0.002	.69	60.04*	0.002	.63
Time Period (1, 35)	0.02	0.001	.00	0.87	0.001	.02
Item Status x Time Period (1, 35)	81.05*	0.000	.70	59.42*	0.000	.63

Note: * = $p < .05$

Table 3

Effects of Level-of-processing (LOP) Item Status on Pupil Dilation Ratio at Test: 2 (LOP: Shallower vs. Deeper) x 2 (Time Period: Word vs. Response) ANOVAs

	Shallow Context			Deep Context		
Effect (*df*)	*F*	*MSE*	η^2	*F*	*MSE*	η^2
LOP (1, 35)	4.81*	0.002	.12	0.04	0.001	.00
Time Period (1, 35)	5.92*	0.002	.15	11.73*	0.001	.25
LOP x Time Period (1, 35)	3.71	0.001	.10	1.80	0.000	.05

Note: * = $p < .05$

Effects of Recollect/Familiar Judgments on the Pupil at Test

Figure 5 provides the mean PDRs at test, this time separated as a function of the judgment type made at test. Analyses of the mean PDRs associated with trials leading to recollect or familiar judgments (Figure 5, Panels C and D) are reported in Table 4, excluding new judgments because they were rarely made for medium and deep items.

Pupil dilation was not significantly greater for recollect judgments than for familiar judgments in either context group (i.e., there were no main effects of judgment). There was a main effect of time period reflecting greater dilation during the response (vs. word) period. More importantly, there was an interaction between judgment and time period in both test contexts, and its form was very different across contexts, as highlighted in Figure 6. In the shallow context, there was a recollect/familiar pupil effect during the word period, $t(34) = 2.56$, $SEM = 0.006$, but no difference during the response period, $t < 1$. In the deep context, there was no

Table 4

Effects of Recollect/Familiar Judgment on Pupil Dilation Ratio at Test: 2 (Judgment: Recollect vs. Familiar) x 2 (Level-of-processing: Shallower vs. Deeper) x 2 (Time Period: Word vs. Response) ANOVAs

	Shallow Context			Deep Context		
Effect (*df*)	*F*	*MSE*	η^2	*F*	*MSE*	η^2
Judgment (1, 35)	0.77	0.003	.02	3.08	0.008	.09
LOP (1, 35)	0.21	0.008	.01	.09	0.004	.00
Time Period (1, 35)	17.41*	0.005	.34	16.68*	0.004	.36
Judgment x LOP (1, 35)	0.28	0.007	.01	0.21	0.003	.01
Judgment x Time Period (1, 35)	8.89*	0.001	.21	22.20*	0.001	.43
LOP x Time Period (1, 35)	0.42	0.002	.01	.89	0.001	.03
Judgment x LOP x Time Period (1, 35)	0.16	0.001	.01	0.52	0.000	.02

Note: * = $p < .05$. There were some missing values thus for recollect/familiar judgments in the shallow context *df* = 1 and 34, and in the deep context *df* = 1 and 30.

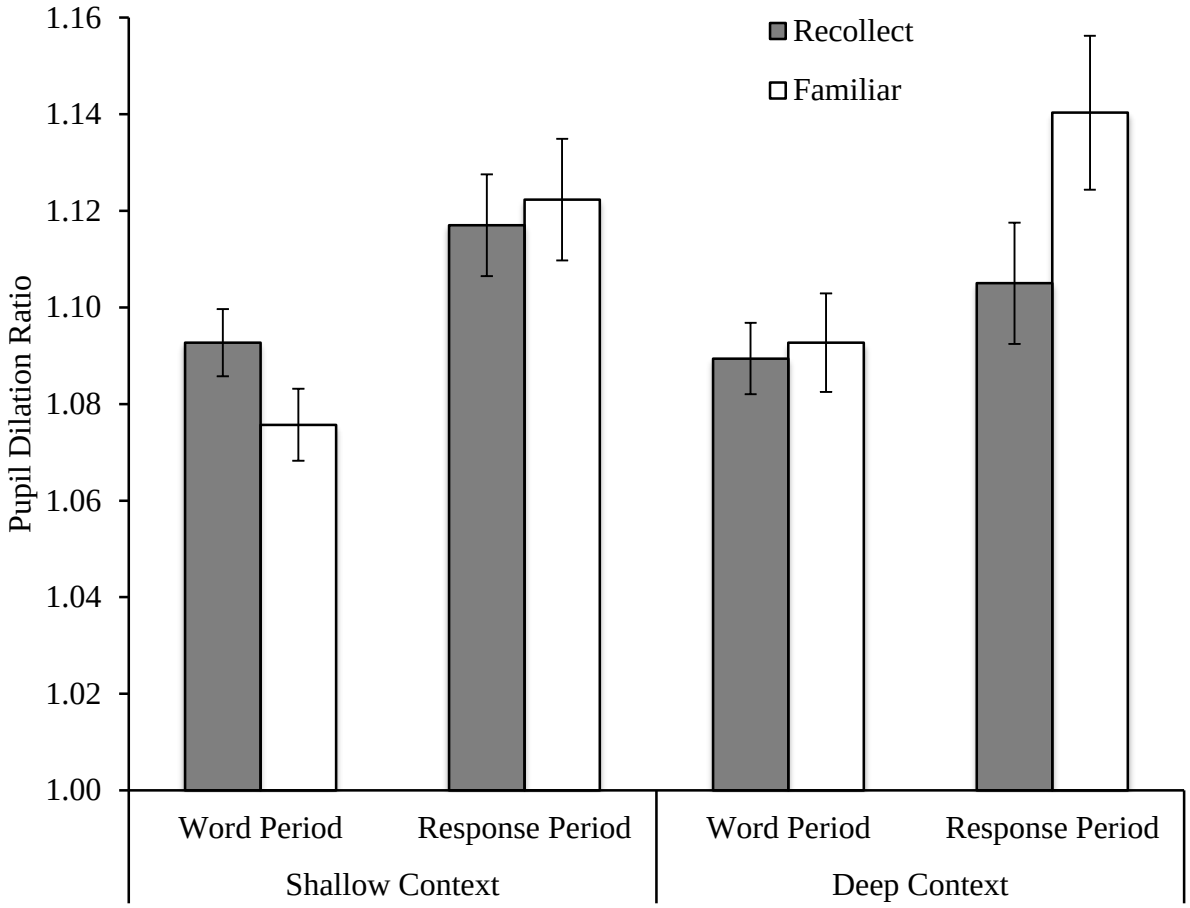

Figure 6. Mean pupil dilation ratios at test for studied items as a function of judgment type and time period. Error bars indicate standard error.

recollect/familiar pupil effect during the word period, $t < 1$—an unexpected result, given Otero et al.'s (2011) finding—but a reverse effect (i.e., smaller PDR for recollect than familiar judgments) during the response period, $t(30) = 2.69$, $SEM = 0.013$. Thus, Otero et al.'s (2011) effect was replicated, but only in the shallow context, and only during the word period. One commonality across contexts was a striking increase in PDR associated with familiar judgments across time periods (see Figure 6). These results are followed up in the discussion.

Also notable was the absence of a LOP effect on the pupil at test as a function of the judgment made. Contrary to Otero et al. (2011) deeper LOP was not associated with greater pupil dilation in either test context (i.e., no main effect of LOP), and this held regardless of judgment (i.e., no interactions between judgment and LOP). These null effects are surprising given the LOP effects based on item status (i.e., ignoring judgment type) reported earlier. Importantly, then, the effects of LOP on recollect/familiar judgments were not mirrored in the pupil, even though the pupil was sensitive to *whether* items were studied (old/new effect) and was sometimes sensitive to *how* items were studied (LOP effect in the shallow context). These findings, discussed more below, will require replication.

I conducted Bayesian analyses (Masson, 2011) comparing the strength of evidence supporting a model that assumes there is no LOP effect on each judgment type (i.e., a null model) against a model that assumes there is an effect. The estimated probability that the null model was preferred was $P_{BIC} = .86$ for recollected items and $P_{BIC} = .85$ for familiar items in the shallow context, and $P_{BIC} = .86$ for recollected items, and $P_{BIC} = .85$ for familiar items in the deep context. This lends some support to the possibility that the pupil does not differentiate between experiences that lead to recollect versus familiar judgments.

In summary, the effects of LOP on the pupil yielded several findings useful in evaluating the basis of pupil responses during recognition. The old/new pupil effect, when response is ignored, confirms that the pupil responds to prior study. The LOP effect in the shallow context further suggests that the pupil can also respond to the depth of encoding. As delineated in the discussion, the LOP effect suggests that during recognition the pupil responds to re-experiencing of encoding and/or to the memory strength elicited by the word at test. Finally, the pattern of pupil effects across judgment types and time periods suggests that the pupil is sometimes (though not always) sensitive to whether recollection or familiarity is experienced.

Effects of Test Context on Recollect/Familiar Judgments at Test

This section reports whether the effects of test context on recognition judgments replicated when study context was controlled (cf. Bodner and Lindsay, 2003; Tousignant & Bodner, 2012; Tousignant et al., 2014). To this end, paired samples t-tests were used to compare recollection judgments, familiarity judgments, and IRK familiarity estimates for medium items across test contexts (see Figure 4). New items were not analyzed because of low false recognition rates (especially for recollection judgments). Medium LOP items were more likely to receive recollect judgments in the shallow (vs. deep) context, $t(70) = 2.18$, $SEM = 0.05$, and were also less likely to receive familiar judgments, $t(70) = 2.61$, $SEM = 0.04$. The IRK estimate of familiarity for medium LOP items (excluding participants with recollect rates of 1 and/or familiar rates of 0) did not differ between the shallow and deep test context (.83 vs. .87), $t < 1$. All three comparisons replicated Bodner and Lindsay (2003) while holding constant the study context. Hence, context effects on recollection are not driven by differences in study context (see also Bodner & Lindsay, Experiment 2).

Effects of Test Context on Pupil Dilation for Medium LOP Items at Test

The results thus far provide new evidence that the pupil can be sensitive to prior study, to depth of encoding, and to differences between recollect and familiarity experiences. I also confirmed that recollect/familiarity judgments are sensitive to test context. Here, I examined whether the test context modulated pupil dilation for the medium LOP items common to both groups. I first examined this question regardless of judgment given at test (Figure 5, Panel A). Mean PDRs were analyzed as laid out in Table 5. Critically, there was no main effect of test context: The pupil's reaction to medium items was similar across test contexts even though rates of recollect and familiar judgments differed. The main effect of time period here again reflected greater dilation during the response (vs. word) period at test.

Next, I examined the critical issue of whether test context influenced the pupil's response to medium LOP items receiving recollect versus familiar judgments (see Figure 5, Panel C and D; see Table 6 for the analyses). There was again no effect of test context. Although the interaction between test context and judgment was marginal (p = .07), the effect of test context did not reach significance either for medium items judged recollect (Table 6, middle) or familiar (Table 6, bottom). The interaction between judgment and time period reflected the earlier result documented in Figure 6.

Medium items received different rates of recollect versus familiar judgments across test contexts, yet the pupil was insensitive to test context. As a result, the pupil pattern could not be used to triangulate whether test context has a crossover, single-dissociation, or parallel effect on recollection versus familiarity (see Figure 2). I conducted Bayesian analyses comparing the strength of evidence supporting the null effect model that assumes no test context effect on the pupil against a model assuming an effect. The probability favoring the null model was P_{BIC} = .88

Table 5

Effects of Test Context on Pupil Dilation Ratio for Medium Items at Test: 2 (Test Context: Shallow vs. Deep) x 2 (Time Period: Word vs. Response) ANOVA

Effect (*df*)	*F*	*MSE*	η^2
Test Context (1, 70)	0.06	0.004	.00
Time Period (1, 70)	21.55*	0.001	.24
Test Context x Time Period (1, 70)	0.02	0.001	.00

Note: * = $p < .05$.

Table 6

Effects of Test Context and Recollect/Familiar Judgment on Pupil Dilation Ratio at Test for Medium Level-of-Processing Items: 2 (Test Context: Shallow vs. Deep) x 2 (Judgment: Recollect vs. Familiar) x 2 (Time Period: Word vs. Response) ANOVA

Judgment/Effect (*df*)	*F*	*MSE*	η^2
Recollect vs. Familiar Judgments			
Test Context (1, 68)	0.68	0.012	.01
Judgment (1, 68)	0.69	0.005	.01
Time Period (1, 68)	32.77*	0.003	.33
Test Context x Judgment (1, 68)	3.42	0.005	.05
Test Context x Time Period (1, 68)	0.21	0.003	.00
Judgment x Time Period (1, 68)	14.85*	0.001	.18
Test Context x Judgment x Time Period (1, 68)	0.11	0.001	.00
Recollect Judgments			
Test Context (1, 70)	0.22	0.005	.00
Time Period (1, 70)	15.36*	0.001	.18
Test Context x Time Period (1, 70)	0.09	0.001	.00
Familiar Judgments			
Test Context (1, 68)	2.08	0.012	.03
Time Period (1, 68)	35.41*	0.003	.34
Test Context x Time Period (1, 68)	0.24	0.003	.00

Note: * = $p < .05$. The *df*s for familiar judgments are lower due to missing values.

for medium items judged recollect, and P_{BIC} = .74 for medium items judged familiar. These analyses lend support to the possibility that the pupil responded similarly to the medium items across test contexts. I return to this important result in the discussion.

In sum, the pupil did not respond to the test context manipulation. The discussion section considers this pattern in light of attributional accounts of test context effects. Because it was not sensitive to context, the pupil was also not useful for triangulating which behavioral measure (binary judgments vs. IRK estimates of familiarity vs. independent ratings) most accurately describes the true effect of test context on recollection and familiarity.

Does Pupil Dilation at Study Predict Recognition at Test?

Lastly, I examined whether pupil dilation at study was predictive of pupil dilation at test, and/or of the recognition judgment made at test. I performed linear regressions at the level of individual trials/items, separately for the word and response time periods. PDR values at study and test were very weakly (but significantly) positively correlated during both the word period, R^2 = .006, $F(1, 2869) = 18.14$, $MSE = 0.011$, and the response period, R^2 = .004, $F(1, 2864) = 11.34$, $MSE = 0.011$. These small correlations could indicate that during recognition, pupil response is weakly influenced by reinstatement of aspects of the study experience.

I also examined whether PDR at study predicted judgment at test. For each type of studied item I computed each subject's mean PDR at study for items that went on to receive either recollect, familiar, or new judgments (see Figure 7; Table 7). The main effect of time period echoes previous results of greater dilation in the word (vs. response) period at study. Pupil dilation at study did not differ based on the judgment at test during either time period. Thus, although pupil dilation at study and at test were weakly correlated, pupil dilation at study was not predictive of the recognition judgment given at test (cf. Naber et al., 2013; Papesh et al., 2012).

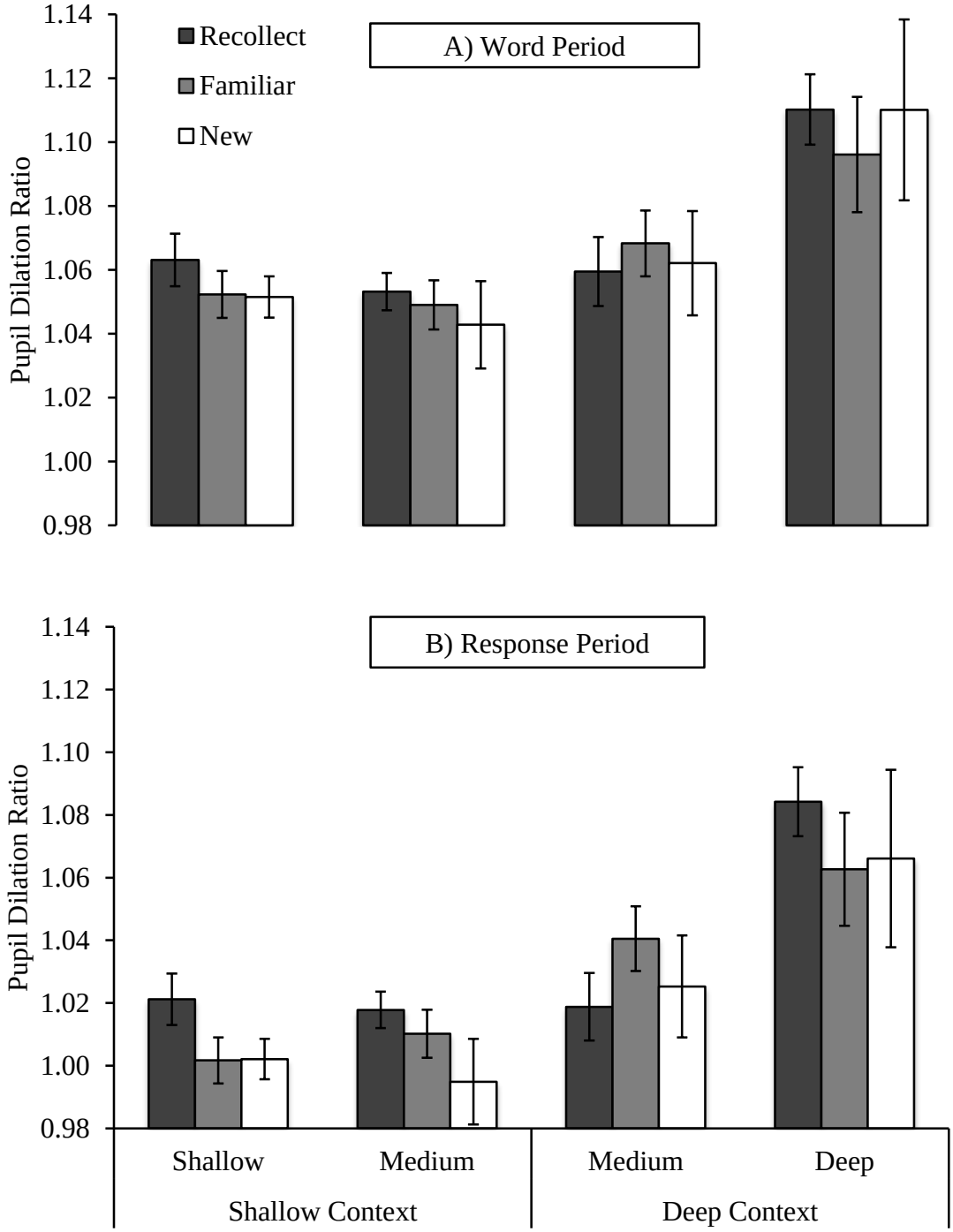

Figure 7. Mean pupil dilation ratios at study as a function of judgment at test and time period. Error bars indicate standard error.

Table 7

Pupil Dilation Ratio at Study as a Function of Judgment at Test: 3 (Judgment at Test: Recollect vs. Familiar vs. New) x 2 (Time Period at Study: Word vs. Response) ANOVAs

Test Context/Level-of-processing(LOP)/Effect (*df*)	*F*	*MSE*	η^2
Shallow Context			
Shallow LOP Items			
Judgment at Test (2, 66)	1.71	0.003	.05
Time Period at Study (1, 33)	50.85*	0.002	.61
Judgment at Test x Time Period at Study (2, 66)	0.53	0.001	.02
Medium LOP Items			
Judgment at Test (2, 36)	0.63	0.004	.03
Time Period at Study (1, 18)	29.69*	0.002	.62
Judgment at Test x Time Period at Study (2, 36)	0.36	0.001	.02
Deep Context			
Medium LOP Items			
Judgment at Test (2, 38)	0.40	0.006	.02
Time Period at Study (1, 19)	17.45*	0.002	.48
Judgment at Test x Time Period at Study (2, 38)	0.41	0.001	.02
Deep LOP Items			
Judgment at Test (2, 30)	0.14	0.018	.01
Time Period at Study (1, 15)	12.05*	0.002	.45
Judgment at Test x Time Period at Study (2, 30)	0.30	0.002	.02

Note: * = $p < .05$. The *df*s for medium and deep items are lower due to missing values.

CHAPTER 4: DISCUSSION

My thesis investigated the effects of test context on the subjective experiences of recollection and familiarity that arise during recognition. The functional account posits that test context influences the definitions of recollection and familiarity such that medium items become more likely to be judged as recollect (vs. familiar) in the shallow (vs. deep) test context (Bodner & Lindsay, 2003). Alternately, medium items may be experienced more strongly in the shallow (vs. deep) test context. Pupil dilation was used to discriminate between these possibilities. To this end, I used an improved version of Bodner and Lindsay's design to clarify what the pupil responds to during recognition. Conversely, I also used the pupil as an independent measure of the effects of LOP and test context on recollection and familiarity. Below I discuss my main findings and their implications, as well as limitations of the study and future research directions.

Isolating the effects of LOP and test-list context on recollect/familiar judgments

I controlled for study duration and production factors that were confounded with the LOP manipulation in prior studies of test context effects (Bodner & Lindsay, 2003; Tousignant & Bodner, 2012; Tousignant et al., 2014). These improvements confirmed that deeper LOP, per se, led to more recollect judgments and fewer familiar judgments but no change in the IRK estimate of familiarity (Yonelinas, 2002). Using a purer manipulation of LOP also allowed me to isolate the effects of LOP on pupil dilation.

I also controlled the study context, which had covaried with the test context in previous studies of test context effects on recognition (Bodner & Lindsay, 2003; Tousignant & Bodner, 2012; Tousignant et al., 2014; but see Bodner & Richardson-Champion, 2007). This improvement confirmed that medium items received more recollect judgments and fewer

familiar judgments in the shallow (vs. deep) test context, and established the role of test context. Using a pure manipulation of test context allowed me to isolate its effects on pupil dilation.

Using LOP and test context to clarify the basis of pupil dilation effects

Having obtained pure effects of LOP and test context, I used these manipulations to help elucidate the bases of pupil dilation at study and test. Deeper LOP tasks led to greater pupil dilation than shallower LOP tasks at study, a novel finding. This pattern is consistent with the assumption that deeper LOP tasks are more cognitively effortful (Otero et al., 2011; see Beatty & Lucero-Wagoner for a review), though it remains possible that the pupil responded to other correlates of the LOP manipulation such as the amount of elaborative processing required or the number of episodic memories elicited by items in each LOP task. Further research should aim to specify the basis of the LOP effect on the pupil at study.

During recognition, I found some evidence that the pupil responds to aspects of encoding as well as to experiences of recollection and familiarity. Three possibilities of what drives pupil dilation during recognition were considered. First, as suggested by Papesh et al. (2012) the pupil may respond to re-experiencing of encoding (see Figure 1, Panel B). Second, as suggested by Otero et al. (2011), dilation may reflect the memory strength experienced for the item at test (Figure 1, Panel C). Third, as suggested by Vo et al. (2008), the pupil may respond to the recognition effort required (Figure 1, Panel D).

The pupil old/new pupil effect is potentially consistent with all three possibilities. Re-experiencing of encoding effort would yield greater pupil dilation for old items. However, the pupil also responds to new items, hence other factors clearly influence dilation at recognition. Greater memory strength at test would also yield greater pupil dilation for old items. Lastly, old items may require more effort to classify (i.e., recognition effort) at test because participants

must evaluate the basis of their recognition experiences, which are more likely to be absent for new items.

To differentiate these three possibilities, I first examined the effect of LOP item status on the pupil at test (ignoring judgment given to items). I found greater dilation for medium (vs. shallow) LOP items, suggesting the pupil does not chiefly respond to the recognition effort (cf. Figure 1, Panel D; see also Otero et al., 2011). On the other hand, greater dilation was not found for deep (vs. medium) items, though I suggest this may be due to ceiling recognition in this context. The effect of LOP item status on the pupil at test thus suggests that during recognition, the pupil primarily responds to re-experiencing of cognitive effort from encoding and/or to the memory strength elicited by the item at test (Figure 1, Panels B-C). To adjudicate between these possibilities, I examined the effects of test context on pupil dilation for medium items.

Test context did not influence the pupil's dilation to medium LOP items (again, ignoring judgment type). This outcome is consistent with the re-experiencing of effort from encoding basis (Figure 1, Panel E), given the study context and hence the encoding efforts for medium items were the same for both groups. Alternatively, medium items may simply be experienced equally strongly in the two contexts, in which case the null effect is also consistent with the memory strength basis (Figure 1, Panel G). The null effect of context on the pupil suggests that test context influences how participants define recollection and familiarity (a decision-based account) rather than the raw strength of their raw recognition experiences (a memory-based account) (see McCabe & Balota, 2007).

Three additional findings suggest that re-experiencing of effort from encoding plays a limited role in driving pupil dilation at test. First, dilation at study and test were only weakly correlated. Second, pupil dilation at study did not differ based on the judgment received at test

(cf. Naber et al., 2013). Third, the pupil also dilated to new items where reinstatement of encoding was not possible, resulting in PDR values greater than 1 (Figure 5, Panel A).

The present results also raise the novel possibility that pupil dilation at test might reflect a form of effort different from recognition effort (cf. Vo et al., 2008). Specifically, the pupil may reflect the effort exerted in distinguishing between recollect and familiar experiences for items that are recognized (i.e., *recollection effort*). The recollect/familiar pupil effect during the word period in the shallow context, and the reverse recollect/familiar pupil effect during the response period in the deep context, both point to a larger increase in pupil dilation across time periods for items judged familiar (vs. recollect). Yonelinas (2002) argued that participants make a familiar judgment only once recollection fails. The recollection attempt likely ceases once participants retrieve a piece of information that allows them to make a recollection judgment. When recollection fails and a familiarity judgment is made, this recollection effort would necessarily have lasted longer. The increase in pupil dilation across time periods for items judged familiar may be indicative of the "nagging" feeling of familiarity that occurs when no information is retrieved to anchor a recognition experience to a specific context (i.e., recollection). Consistent with this possibility, recognition judgments based on familiarity (vs. recollection) typically take longer (Dewhurst, Holmes, Brandt, & Dean, 2006; Duarte, Ranganath, Winward, Hayward, & Knight, 2004; Wixted & Stretch, 2004).

The possibility that the pupil is sensitive to recollection effort is consistent with two of my other findings. First, pupil dilation was greater during the response (vs. word) period for studied items, but vice versa for new items (see Figure 5, Panel A). For old items, recollection effort would typically increase across time periods, as participants work to resolve the origin of their recognition experience. In contrast, for new items, the initial recognition experience would

typically be weak or absent and thus the recollection effort might be stopped before the response period. Second, the recollection effort notion is consistent with the effects of LOP item status on pupil dilation (Table 3). The initial recognition experience would typically be richer for deeper LOP items, spurring a greater need to engage in recollection effort to justify a recollect judgment.

In sum, the pupil appears to respond to more than one factor at test, including recollection effort and memory strength. The pupil's response to an item at test is also at least weakly related to its response to the same item at study. I now turn to a consideration of how the pupil informs us about the basis of test context effects on recollection and familiarity.

Using pupil dilation effects to clarify the basis of test context effects on recognition judgments

Bodner and Lindsay (2003) suggested that the dissociative effects of test context on recognition judgments were consistent with Grupposo et al.'s (2007) functional account. According to the functional account, participants' definitions of subjective recognition experiences are influenced by task and context factors, reflected in different rates of recollection and familiarity judgments for medium items across contexts. Alternately, however, medium items may have been experienced more strongly in the shallow (vs. deep) test context, resulting in more recollection judgments. Assuming that the pupil responds to memory strength at test (Montefinese et al., 2013; Otero et al., 2011), the absence of test context differences on pupil dilation to medium LOP items suggests they were experienced equally strongly in either context. Thus, the behavioural dissociation appears to be due to changing definitions of recollection and familiarity, thus specifying a decision-based rather than memory-based version of the functional account (see also McCabe & Balota, 2007).

It was hoped that pupil dilation would help triangulate which behavioral measure (binary judgments vs. IRK-corrected estimates of familiarity vs. independent ratings) most accurately captures the pattern of test context effects on recognition judgments. Test context exerts a similar effect across these measures on recollection (but see Tousignant et al., 2014 for an exception), but exerts different outcomes for familiarity (see Figure 2). Unfortunately, the pupil was not sensitive to context either for recollect or familiar judgments, and was thus not helpful for adjudicating whether medium items were less familiar, equally familiar, or more familiar in the shallow context than in the deep context. Below this is offered as a subject for future research.

Implications for dual-process and single-process models of memory

In this section, I reconcile my pupil effects with single- and dual-process models of recognition memory. My finding that pupil dilation showed opposite patterns during the word period in the shallow context (recollect > familiar) versus the response period in the deep context (familiar > recollect) suggests that the time courses of the two experiences differ (see Figure 6). This pattern is consistent with recollection and familiarity being independent processes, whereas single-process signal-detection models cannot capture (or explain) such time course differences. Indeed, the reverse recollect/familiar pupil effect implies that familiar items are experienced more strongly than recollected items during the response period, which contradicts the single-process account.

Single-process models are also challenged by the absence of a consistent pupil recollect/familiar effect (cf. Otero et al., 2011). Proponents of these models could suggest that the pupil is just not sensitive to memory strength, yet that possibility is contradicted by the pupil old/new effect as well as by the effect of LOP on pupil dilation (present study, and Otero et al.). Overall, then, the present pupil results appear to mesh better with dual-process accounts.

Limitations and future directions

Several limitations in the present study will need to be addressed in future research, particularly methodological changes that would help clarify the pupil effects. One question is whether the absence of LOP effects on pupil dilation in the deep context group reflected ceiling recognition (despite the LOP effect on recollect judgments). Implementing a study-test delay would resolve this issue, and would also increase recollect/familiar judgments false alarms thus allowing signal-detection analyses to be applied. Signal-detection analyses would be useful for determining whether discrimination for medium items was greater in the shallow context. Such a result would suggest that definitions of recollection can differ in how useful they are for distinguishing these items from new items (Bodner & Lindsay, 2003); cf. Tousignant & Bodner, 2012).

Higher false alarm rates would also enable a comparison of pupil dilation for new items as a function of the judgment made at test, which could not be done in the current study due to low rates of false recognition. Examining pupil dilation for new items would isolate what the pupil responds to at recognition aside from re-experiencing of encoding. The elimination of this factor (no re-experiencing of encoding can occur for new items) would allow examination of other factors driving pupil dilation at test (e.g., perceived memory strength and/or effort of retrieving information from study).

There is also a potential power issue in trying to detect effects of LOP on pupil dilation for recollect versus familiar judgments (see Figure 4). Specifically, deep LOP items often receive recollect judgments, leaving fewer pupil observations for familiar judgments, and vice versa for shallow LOP items. The use of longer study lists, providing more items at test, might help with this issue.

A potential limitation of the test design was my use of a two-step procedure (old/new judgment then recollect/familiar judgment). The two-step procedure is not ideal for comparing pupil dilation for new judgments versus recollect/familiarity judgments because the latter judgments necessarily occur later. Instead, a one-step procedure (recollect/familiar/new judgment), as used by Bodner and Lindsay (2003), may well be preferable.

Another limitation was that luminance differences across the time periods (when the word was replaced with response cues) were confounded with the effects of time course of pupil dilation. These changes in luminance are relevant given that pupil dilation changes due to light reflex are larger than non-reflexive changes due to the cognitive task (Beatty & Lucero-Wagoner, 2000). In addition, the positioning the response cues likely led participants to move their eyes off fixation during the response period. Given that pupil measurement can be affected by pupil position on the screen (Jiye, 2006), this was not idea.

To control for luminance and eye movement throughout the trial, the word should be presented throughout the duration of the trial at study and at test even while participants make their responses. Alternatively, a response deadline could be used such that the target is replaced with a row of Xs after a set duration, and participants are asked to complete their responses before the Xs appear. Such changes would control for luminance and the position of the pupil throughout the whole trial. Finally, executing a response can influence pupil dilation (e.g., Moresi et al., 2008; Van der Molen, Boomsma, Jennings, & Nieuwboer, 1989), thus another change worth considering would be to have participants make their decision mentally but to not respond until after the trial (see van Rijn et al., 2012).

Conclusion

My thesis aimed to use the pupil to discern the basis of test context effects on recollection and familiarity, while contributing to our understanding of the effects of recognition memory on pupil dilation. To these ends, using pure manipulations of LOP and test context I clarified that during recognition, the pupil responds to depth of encoding, as well as to experiences of recollection and familiarity. Finding no effects of test context on the pupil, I also concluded that test context influences the definitions of recollection and familiarity that participants use rather than the strength of their recognition experiences.

References

Ariel, R., & Castel, A. D. (2014). Eyes wide open: Enhanced pupil dilation when selectively studying important information. *Experimental Brain Research, 232*, 337-344.

Beatty, J., & Kahneman, D. (1966). Pupillary changes in two memory tasks. *Psychonomic Science, 5,* 371-372.

Beatty, J., & Lucero-Wagoner, B. (2000). The pupillary system. in J. T. Cacioppo, L. G. Tassinary, & G. G. Berntson (Eds.) *Handbook of psychophysiology* (2nd ed., pp. 142-162). New York, NY: Cambridge University Press.

Bodner, G. E., & Lindsay, D. S. (2003). Remembering and knowing in context. *Journal of Memory and Language, 48*, 563-580.

Bodner, G. E., & Richardson-Champion, D. D. L. (2007). Remembering is in the details: Effects of test-list context on memory for an event. *Memory, 15*, 718-729.

Bradshaw, J. L. (1969). Background light intensity and the pupillary response in a reaction time task. *Psychonomic Science, 14*, 271-272.

Brown, A. A., & Bodner, G. E. (2011). Re-examining dissociations between remembering and knowing: Binary judgments vs. independent ratings. *Journal of Memory and Language, 65*, 98-108.

Clark, W. R., & Johnson, D. A. (1970). Effects of instructional set on pupillary response during a short-term memory task. *Journal of Experimental Psychology, 85,* 315-317.

Dewhurst, S. A., Holmes, S. J., Brandt, K. R., & Dean, G. M. (2006). Measuring the speed of the conscious components of recognition memory: Remembering is faster than knowing. *Consciousness and Cognition, 15*, 147-162.

Duarte, A., Ranganath, C., Winward, L., Hayward, D., & Knight, R. T. (2004). Dissociable neural correlates for familiarity and recollection during encoding and retrieval of pictures. *Cognitive Brain Research, 18,* 255-272.

Donaldson, W. (1996). The role of decision processes in remembering and knowing. *Memory & Cognition, 24,* 523–533.

Dunn, J. C. (2004). Remember Know: A matter of confidence. *Psychological Review, 111,* 524-542.

Gardiner, J. M. (1988). Functional aspects of recollective experience. *Memory & Cognition, 16,* 309–313.

Gardiner, J. M., & Richardson-Klavehn, A. (2000). Remembering and knowing. In E. Tulving & F. I. M. Craik (Eds.), *Handbook of memory* (pp. 229-244). New York: Oxford University Press.

Gardner, R. M., Mo, S. S., & Borrego, R. (1974). Inhibition of pupillary orienting reflex by novelty in conjunction with recognition memory. *Bulletin of the Psychonomic Society, 3,* 237-238.

Gardner, R. M., Mo, S. S., & Krinsky, R. (1974). Inhibition of pupillary orienting reflex by heteromodal novelty. *Bulletin of the Psychonomic Society, 4,* 510-512.

Goldinger, S. D., & Papesh, M. H. (2012). Pupil dilation reflects the creation and retrieval of memories. *Current Directions in Psychological Science, 21,* 90-95.

Granholm, E., Asarnow, R. F., Sarkin, A. J., & Dykes, K. L. (1996). Pupillary responses index cognitive resource limitations. *Psychophysiology, 33,* 457-461.

Gruppuso, V., Linsday. D. S., & Kelley, C. M. (1997). The process-dissociated procedure and similarity: Defining and estimating recollection and familiarity in recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 23,* 259-278.

Heaver, B., & Hutton, S. B. (2011). Keeping an eye on the truth? pupil size changes associated with recognition memory. *Memory, 19,* 398-405.

Hess, E. H., & Polt, J. M. (1960). Pupil size as related to interest value of visual stimuli. *Science, 132,* 349-350.

Hess, E. H., & Polt, J. M. (1964). Pupil size in relation to mental activity during sample problem solving. *Science, 143,* 1190-1192.

Higham, P. A., & Vokey, J. R. (2004). Illusory recollection and dual-process models of recognition memory. *The Quarterly Journal of Experimental Psychology A: Human Experimental Psychology, 57,* 714-744.

Jacoby, L. L. (1991). A process dissociation framework: Separating automatic from intentional uses of memory. *Journal of Memory and Language, 30,* 513-541.

Jacoby, L. L., & Dallas, M. (1981). On the relationship between autobiographical memory and perceptual learning. *Journal of Experimental Psychology: General, 110,* 306-340.

Jainta, S., & Baccino, T. (2010). Analyzing the pupil response due to increased cognitive demand: An independent component analysis study. *International Journal of Psychophysiology, 77,* 1-7.

jiye@sr-research.com (2006, March 27). Pupil size values [Msg 2]. Message posted to https://www.sr-support.com/showthread.php?218-Pupil-size-values

Kahneman, D., & Beatty, J. (1966). Pupil diameter and load on memory. *Science, 154,* 1583-1585.

Kahneman, D., & Peavler, W. S. (1969). Incentive effects and pupillary changes in association learning. *Journal of Experimental Psychology, 79,* 312-318.

Kuchinke, L., Vo, Melissa, 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.

Kurilla, B. P., & Westerman, D. L. (2008). Processing fluency affects subjective claims of recollection. *Memory & Cognition, 36,* 82-92.

MacLeod, C. M., Gopie, N., Hourihan, K. L., Neary, K. R., & Ozubko, J. D. (2010). The production effect: Delineation of a phenomenon. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 36,* 671-685.

McCabe, D. P., & Balota, D. A. (2007). Context effects on remembering and knowing: The expectancy heuristic. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 33*(3), 536-549.

Masson, E. J. M. (2011). A tutorial on a practical Bayesian alternative to null-hypothesis testing. *Behavioral Research, 43,*

Montefinese, M., Ambrosini, E., Fairfield, B., & Mammarella, N. (2013). The subjective pupil old/new effect: Is the truth plain to see? *International Journal of Psychophysiology, 89,* 48-56.

Moresi, S., Adam, J. J., Rijcken, J., Van Gerven, P. W. M., Kuipers, H., & Jolles, J. (2008). Pupil dilation in response preparation. *International Journal of Psychophysiology, 67,* 124-130.

Naber, M., Frassle, S., Rutishauser, U., Einhauser, W. (2013). Pupil size predicts novelty and predicts later retrieval success for declarative memories of natural scenes. *Journal of Vision, 13,* 1-20.

Nuthmann, A., & van der Meer, E. (2005). Time's arrow and pupillary response. *Psychophysiology, 42,* 306-317.

Otero, S. C., Weekes, B. S., & Hutton, S. B. (2011). Pupil size changes during recognition memory. *Psychophysiology, 48*(10), 1346-1353.

Papesh, M., H., Goldinger, S. D., & Hout, M. C. (2012). Memory strength and specificity revealed by pupillometry. *International Journal of Psychophysiology, 83,* 56-64.

Rugg, M. D., Mark, R. E., Walla, P., Schloerscheidt, A. M., Birch, C. S., & Allan, K. (1998). Dissociation of the neural correlates of implicit and explicit memory. *Nature, 392,* 595-598.

Schluroff, M., Zimmermann, T. E., Freeman, R. B. Jr., Hofmeister, K., Lorscheid, T., & Webber, A. (1986). Pupillary responses to syntactic ambiguity of sentences. *Brain and Language, 27,* 322-344.

Tousignant, C., & Bodner, G. E. (2012). Test context affects recollection and familiarity ratings: Implications for measuring recognition experiences. *Consciousness and Cognition: An International Journal, 21*, 994-1000.

Tousignant, C., Bodner, G. E., & Arnold, M. (2014). *Effects of context on recollection and familiarity experiences are task dependent*. Manuscript in preparation.

Tousignant, C., Mulji, R., & Bodner, G. E. (2008). *Test-list context affects remember judgment discrimination*. Unpublished manuscript.

Tulving (1985). Memory and consciousness. *Canadian Psychologist, 26,* 1-12.

Tulving, E., & Markowitsch, H. J. (1998). Episodic and declarative memory: Role of the hippocampus. *Hippocampus, 8,* 198-204.

Van der Molen, M. W., Boomsma, D. I., Richard, J. J., & Nieuwboer, R. T. (1989). Does the heart know what the eye sees? A cardiac/pupillometric analysis of motor preparation and response execution. *Psychophysiology, 26,* 70-80.

van Rijn, H., Dalenberg, J. R., Borst, J. P., Sprenger, S. A. (2012). Pupil dilation co-varies with memory strength of individual traces in a delayed response paired-associate task. *PLoS ONE, 7,* e51134.

Võ, M. L.-H., Jacobs, A. M., Kuchinke, L., Hofmann, M., Conrad, M., Schacht, A., & Hutzler, F. (2008). The coupling of emotion and cognition in the eye: Introducing the pupil old/new effect. *Psychophysiology, 45,* 130–140.

Whittlesea, B. W. A. (2002a). False memory and the discrepancy-attribution hypothesis: The prototype-familiarity illusion. *Journal of Experimental Psychology: General, 131,* 96-115.

Whittlesea, B. W. A. (2002b). Two routes to remembering (and another to remembering NOT). *Journal of Experimental Psychology: General, 131,* 325-348.

Whittlesea, B. W. A., & Leboe, J. P. (2000). The heuristic basis of remembering and classification: fluency, generation, and resemblance. *Journal of Experimental Psychology: General, 129,* 48-106.

Whittlesea, B. W. A., & Wright, R. L. (1997). Implicit (and explicit) learning: Acting adaptively without knowing the consequences. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 23,* 181-200.

Wiese, H., & Daum, I. (2006). Frontal positivity discriminates true from false recognition. *Brain Research, 1075,* 183-192.

Wixted, J. T., & Stretch, V. (2004). In defense of the signal detection interpretation of remember/know judgements. *Psychonomic Bulletin & Review, 11,* 616-641.

Yonelinas, A. P. (2002). The nature of recollection and familiarity: A review of 30 years of research. *Journal of Memory and Language, 46,* 441-517.

Yonelinas, A. P., & Jacoby, L. L. (1995). The relation between remember and knowing as bases for recognition: Effects of size congruency. *Journal of Memory and Language, 34,* 622-643.

Yonelinas, A. P., Kroll, N. E. A., Dobbins, I., Lazzara, M., & Knight, R. T. (1998). Recollection and familiarity deficits in amnesia: Convergence of remember-know, process dissociation, and receiver operating characteristic data. *Neuropsychology, 12,* 323-339.