Your Effort is Showing!

Pupil Dilation Reveals Memory Heuristics

Megan H. Papesh Stephen D. Goldinger

Arizona State University

Correspondence:

Stephen D. Goldinger
Department of Psychology
Box 871104
Arizona State University
Tempe, Arizona 85287-1104

phone: (480) 965-0127
fax: (480) 965-8544
e-mail: goldinger@asu.edu

It is an honor to contribute to a collection of essays celebrating Bruce Whittlesea's career. The research and ideas from Whittlesea and his colleagues have heavily influenced much of the research in our laboratory, particularly our studies of face perception and memory. Although face processing is often considered "modular" (i.e., highly specialized in neural and computational terms; Haxby, Hoffman, & Gobbini, 2000), we have consistently observed that judgments of face memory are affected by the evaluative and heuristic processes that Whittlesea has hypothesized (e.g. Whittlesea & Leboe, 2000). In this chapter, we briefly review several prior findings that connect Whittlesea's (1997) SCAPE framework with face memory. We then describe new results, wherein we hypothesize that long-term struggles from the SCAPE evaluation system may inspire a new heuristic (kindly dubbed the *"oh… screw it"* heuristic).

As Whittlesea has argued, in any memory test, people are influenced by multiple sources of information. Consider a study by Jacoby and Whitehouse (1989), wherein people were shown words for memorization. In a later recognition test, words were preceded by identity primes that were either masked (subliminal), or were clearly shown. Given subliminal matching primes, the frequency of "old" responses, especially false alarms, increased. Conversely, with overt matching primes, "old" responses decreased. Jacoby and Whitehouse proposed that, with subliminal primes, people interpret enhanced perceptual fluency as familiarity. With overt primes, people discount "positive" signals (either fluency or true familiarity) as a natural by-product of the priming words.

Building upon such findings, Whittlesea and Leboe (2000; Whittlesea & Williams, 1998; 2001) suggested that recognition entails two stages: First is *production* of mental states, wherein images or ideas are brought to mind. Production may follow perceptual input, such as a face, which the mind immediately elaborates (Neisser, 1967), perhaps with a name. Following production, the second stage is *evaluation*. This is not direct stimulus evaluation, such as deciding whether a recognition target exceeds criterion. Instead, Whittlesea (1997; Leboe & Whittlesea, 2002) suggested that people automatically evaluate their own production functions, keeping a running index of the relative harmony of mind. By its nature, evaluation is based on subjective states of mind. For example, imagine encountering a co-worker, whom you easily recognize. On most days, such an encounter will produce immediate recognition, and the evaluation process will not be unduly aroused, creating no particular feelings. Imagine, however, that you encounter the co-worker after he has shaved his long-standing mustache. The production process easily recognizes your acquaintance, but with an uncomfortable dysfluency – something is different and weird. This momentary processing hitch motivates a careful search, as you try to figure out "what's different?"

According to SCAPE, such feelings arise from a *discrepancy-attribution* process (Whittlesea & Williams, 2001). Depending upon context, people have different implicit expectations of processing fluency. When those expectations are violated, an evaluation "flag" is raised, automatically triggering a search for some explanation. Sometimes, the context itself provides a natural attribution – in Jacoby and Whitehouse (1989), inexplicable perceptual fluency evoked feelings of familiarity. In a challenging recognition test, momentary boosts in fluency should evoke the obvious attribution of prior experience. Counter to intuition, when processing expectations are violated, people experience feelings of familiarity. Despite standard usage, truly familiar stimuli generally evoke no feelings of memory. Conversely, mildly familiar stimuli *("what's different about this guy?")* create a strong, nagging sense of familiarity.

In a prior study, Goldinger and Hansen (2005) tested this framework by presenting people with subtle bodily cues that could be experienced as "familiarity signals." They fitted a chair with wireless speakers and participants (sitting in that very chair) memorized words, pictures, and faces (in separate blocks). During recognition, half the test items (old and new) were presented with a simultaneous, subliminal buzz (a low-amplitude, 60-Hz sinewave). In a control condition, the buzz was easy to perceive. People made “old-new” decisions and confidence ratings. To help contrast recollection and familiarity, Goldinger and Hansen presented items that were relatively “easy” and “hard” to remember, such as photos of celebrities and medical students, respectively. When faces were hard to recall, there were clear effects – given the subliminal buzz, people were more likely to respond "old," increasing hits and false-alarms. In the confidence data, when people committed false alarms, the subliminal buzz elicited relatively high confidence. Given hits, the subliminal buzz had the opposite effect, *reducing* confidence. This seemed to reflect the sense of familiarity – when a person really has no memory (for new faces), the buzz created a “tingle” of familiarity and people responded “old” with confidence. But, given true memory (for old faces), the buzz created a tingle of doubt. These findings followed predictions from SCAPE, as the same signal created different memorial interpretations, based on context (Whittlesea & Williams, 2001).

Heuristics in Face Recognition

In SCAPE, the central hypothesis is that memory decisions are often evaluative in nature, as people interpret familiarity cues in context. Lacking absolute criteria for recognition, people rely on memory decision heuristics. Whittlesea and Leboe (2000) described three heuristics, called *generation*, *resemblance*, and *fluency*. In a study of face perception, Kleider and Goldinger (2004) tested the fluency heuristic, as described by Jacoby and Dallas (1981; Jacoby, Kelley, & Dywan, 1989). According to the fluency heuristic, in challenging recognition tasks, people often use the fluency (ease) of perceptual processing as a memory cue. Many data suggest that perceptual processing is enhanced for familiar stimuli: People seem to implicitly assume this relationship, such that “memory illusions” can be elicited by increasing perceptual fluency. As perception is made more fluent, feelings of familiarity arise, leading to increased “old” judgments (a liberal criterion shift).

Although this effect occurs among old items, it is generally larger for new items, because familiarity is their only available cue.

Although face processing is typically robust (Farah, Wilson, Drain, & Tanaka, 1998), Kleider and Goldinger (2004) predicted that fluency manipulations in face perception would affect memory, just as Whittlesea and Leboe (2000) observed with verbal materials. In a series of five experiments, people briefly studied clear faces, followed by a distracter task and a test phase. During the recognition tests, new and old faces were embedded in varying levels of visual white noise. In every experiment, ‘‘old’’ responses increased to clear faces, driven mainly by liberal bias shifts. This pattern occurred regardless of variations in noise levels, warnings to participants that clarity was not related to “old-new” status, and other factors. In two final experiments, Kleider and Goldinger (2004) reversed the pattern: People perceived familiar photos as having greater clarity in noise, and longer presentation durations in a speeded perceptual task. As predicted by SCAPE, increased perceptual fluency created feelings of familiarity, and familiarity created feelings of fluency.

According to Whittlesea and Leboe (2000), memory heuristics fall into two general categories, *quality-of-processing* and *information*. Quality-of-processing heuristics are based on the speed, coherence, and vividness of production – perceptual fluency falls under this domain. Information heuristics are based on contextual or statistical cues. The *generation* heuristic is based on retrieval of episodic details, mentally “placing” stimuli in decision-relevant contexts. When contextual details are highly available, people feel more confident about memory. The *resemblance* heuristic is based on the “fit” between a stimulus under consideration and relevant prior experiences, as when people falsely recognize words that are consistent with a prototype (Roediger & McDermott, 1995).

Moving beyond the fluency heuristic, Kleider and Goldinger (2006) investigated the generation and resemblance heuristics, with special attention to the *other-race effect* (ORE) in face recognition.

Although people are extraordinary face processors (Bahrick, Bahrick, & Wittlinger, 1975), research has consistently shown that people are better able to distinguish among faces from their own race, relative to members of other races (Meissner & Brigham, 2001). For several reasons (see Whittlesea & Leboe, 2000), the generation and resemblance heuristics are difficult to empirically dissociate. Therefore, Kleider and Goldinger (2006) attempted to manipulate the *likely* influence of each heuristic across experiments. Participants were asked to study two series of faces – across conditions, the face sets either combined White and Asian faces, or these subsets were presented in separate lists. In the later recognition task, people were asked to respond "old" only to faces from the second study list. In mixed conditions, people relied heavily on generation (this was indicated by the relative lack of a "resemblance pattern," as people seemed to carefully generate prior list contexts for each face). In the blocked conditions, they relied heavily on resemblance (as evidenced by false-alarm patterns). Once those endpoint conditions were recorded, Kleider and Goldinger (2006) explored the middle ground, "nudging" people back and forth between heuristics by changing the statistical benefits of either strategy.

Eye-Tracking and Pupillometry: Are the Eyes a Window on Memory Processes?

As the foregoing review implies, the basic predictions of SCAPE are easily tested (and have been generally verified) in the domain of face memory. As a rule, virtually all such studies have focused on heuristic processes arising during memory *testing*. It seems completely plausible, however, that people might also adopt heuristic information-processing strategies during *learning*, in preparation for upcoming memory tests. Recall that SCAPE is based on two processes: First is the *production* of mental states, such as seeing a face and trying to create a "strong" memory trace. The second stage is *evaluation* of production efficiency, such as a person feeling confident that she is successfully memorizing the faces. In the case of cross-race faces, however, it seems very likely that self-evaluation will result in decreased confidence, and feelings that future memory will be poor. As such, monitoring behavior (overt and unconscious) during learning trials should provide a window into the evaluation stage of SCAPE.

We have recently begun to examine cognitive phenomena, and the second stage of SCAPE, via psychophysiological indices, specifically the pupillary reflex. It has long been known that, when people perform challenging tasks that require more cognitive effort, their pupils dilate (Porter, Troscianko, & Gilchrist, 2007), possibly representing a summed index of brain activity associated with such tasks (Beatty & Kahneman, 1966). Beatty (1982) summarized the attractive qualities of the pupil reflex that made it Kahneman's (1973) primary index of mental processing load in his theory of attention allocation. As noted by Beatty (1982, p. 276), Kahneman discussed three criteria which should be met by physiological indicators of mental processing load:

> "It should be sensitive to within-task variations in task demands produced by changes in task parameters; it should reflect between-task differences in processing load elicited by qualitatively different cognitive operations; finally, it should capture between-individual differences in processing load as individuals of different abilities perform a fixed set of cognitive operations."

Beatty (1982) reviewed the evidence and concluded that the pupillary reflex satisfies all three of Kahneman's criteria. In our research, we propose that, when used in concert with the appropriate behavioral indices, pupil dilation can serve as a sensitive indicator of underlying cognitive effort, particularly the evaluation of production efficiency, as in SCAPE. The appeal of this often-overlooked dependent measure lies in its independence – the pupil reflex is not subject to task-specific strategies; it is a bias-free measure of processing load that can estimate mental effort expended in a variety of tasks. Next, we discuss several experiments in which we have successfully applied pupillometry to the study of cognitive processing. To preface, across several domains, including face perception, memory, and word naming, we find tight connections between mental effort and pupillary reflexes.

Before turning to a discussion of our research, we must note that, although pupillometry is a sensitive measure of cognitive effort, pupils also change reflexively, based on visual input. Variations in luminance across stimuli, sudden onsets of stimuli, and variations in color can all induce pupillary responses (Porter et al., 2007), necessitating tight experimental control. Porter and Troscianko (2003) discussed several methodological approaches that can minimize unwanted pupil reflexes. These include using relatively low stimulus contrast, avoiding colored stimuli, and using relatively long exposure durations. The use of long exposure durations is particularly germane, as previous research has indicated that task-evoked pupil dilations are a relatively late-arriving index of cognitive effort, beginning several hundred milliseconds following stimulus onset (Kuchinke et al., 2007). In the experiments reported below, we applied a combination of methods to minimize the influence of visually-influenced pupillary reflexes.

Our first investigation examining the nature of the relationship between the pupillary reflex and cognitive effort was another study of the own-race effect (ORE) in face recognition. A number of theories have been proposed to account for the ORE (Levin, 1996; 2000; Maclin & Malpass, 2001; Ng & Lindsay, 1994; Sporer, 2001), but the mechanism by which it occurs is often neglected. We were interested, therefore, in the process by which faces are learned for future recognition. Are there observable encoding differences (across own- and cross-race faces) that could predict future recognition accuracy? That is, can we pinpoint the evaluation stage by examining study trials? We hypothesized that, using eye movements (cf. Henderson, Williams, & Falk, 2005) and pupillometry, we could index the amount of effort expended during learning and relate this to behavioral performance on a recognition memory test.

To investigate the relationship between effort during learning and later recognition accuracy, we presented a group of Asian and Caucasian participants with a set standardized, neutral expression photographs from Ekman & Matsumoto (1993). Critically, all faces were presented in grayscale and for relatively long (5 versus 10 seconds) periods of time. In brief, we found that participants (in both racial groups) who demonstrated an ORE in recognition accuracy selectively *withdrew effort* during the encoding of other-race faces, reflected by both decreased eye movements and pupil dilations (see Goldinger, Yi, & Papesh, 2009, for a full account). Interestingly, we observed a pattern that we compared to "learned helplessness," wherein low-scoring participants' physiological indices of effort steadily declined across trials, essentially indicating that they had succumbed to the newly-dubbed "oh… screw it*"* heuristic.

To investigate, and potentially increase, this effort reduction pattern, we conducted a second experiment, using only Caucasian participants and 10-second exposure durations. Our goal in this experiment was to assess pupil dilation to the exact same faces, but to increase the *perceived difficulty* of the study task by changing the study-list context. Specifically, we added photographs (also from Ekman & Matsumoto, 1993) to the study session, showing the same Asian and Caucasian faces, now with a mixture of different emotional expressions. We expected that, given a longer study phase including emotional faces, people would find the neutral faces progressively less distinctive and interesting. Following SCAPE, we predicted that such perceived difficulty would affect the evaluation stage, such that people would reduce confidence in future memory and selectively withdraw effort from the Asian neutral faces.

Twenty Caucasian students from Arizona State University with normal or corrected-to-normal vision participated for partial course credit. The study stimuli consisted of 52 faces, with equal representation across all variables (sex, emotionality, and race) and no repeated models. All pictures were set to equal mean luminance and were embedded in a black background (1024 x 768) for presentation on a Tobii 1750 17-inch monitor. A chin rest maintained the participants' viewing distance at 60 cm and both eyes were continuously tracked at 50 Hz throughout the experiment. Participants were given 10 seconds to study each face and were later tested on their recognition memory for the photographs using only neutral expression models.

As in our previous experiment, we observed a robust ORE in recognition accuracy, with higher Pr[1] scores to Caucasian faces, $F(1, 19) = 46.06$, MSe = .04, $\eta_p^2 = .45$, and clear differences in eye movement patterns depending on the race of the stimulus face, with participants moving their eyes greater distances across Caucasian faces, $F(1, 19) = 145.8$, MSe = 44.16, $\eta_p^2 = .52$. More relevant to SCAPE were the pupil dilation analyses. Pupil diameters were measured between trials (and while the fixation cross was shown) to establish baseline estimates, then during photograph viewing for comparison. We removed missing observations due to blinks or signal loss, filing those gaps by linear interpolation.[2] Another 0.4% observations were replaced, in the same manner, for values falling more than 2.5 standard deviations from their 10 immediate neighbors. For each participant, we selected the "better" eye (i.e., with fewer corrected observations) for analysis. In an overall analysis of the experiment, we observed significantly greater dilation to Asian faces, relative to Caucasian faces, $F(1, 19) = 185.13$, MSe = 39.99, $\eta_p^2 = .67$, suggesting that Asian faces required greater effort during encoding.

Additional analyses on the eye movement data from this experiment demonstrated that, over the course of the experiment, both high- and low-scoring participants exerted diminishing effort to the encoding of Asian faces. We previously found this pattern in only low-scoring participants. Similar patterns were observed for pupil dilations. As shown in Figure 1, average dilation to Caucasian faces was statistically equivalent across both high- and low-scorers throughout the entire 13-trial encoding period. By contrast, average dilation to the Asian faces declined across trials, with a greater decline

[1] Pr is a common accuracy score, representing the difference between hits and false alarms (see Feenan & Snodgrass, 1990).
[2] This resulted in less than 4.5% of data repair for all participants.

among the low-scoring participants. In an omnibus 2 (Group: high/low) x 2 (Race: Asian/Caucasian) x 13 (Trial Number) ANOVA, we observed a large main effect of Race, $F(1, 19) = 159.6$, MSe = 42.08, $\eta_p^2 = .63$, with greater overall dilation to Asian faces, relative to Caucasian faces. In one-way (Trial Number) ANOVAs on the Asian learning trials for each group, we observed a main effect of Trial Number for both the high-scoring group, $F(1, 9) = 8.47$, MSE = 41.50, $\eta_p^2 = .10$, and the low-scoring group, $F(1, 9) = 23.83$, MSe = 47.18, $\eta_p^2 = .19$. Although the "oh… screw it" pattern was stronger in the low-scoring group, it was now observed in both groups.

Our data suggest that cross-race faces demand extra encoding effort, and that this extra effort manifests itself in differences in pupil diameter and patterns of eye movements. For the present purposes, we focus on the observed differences in pupil dilation, as we believe that they accurately index underlying cognitive effort and predict future memory. Using pupillometry, we are able to investigate the evaluation stage of Whittlesea's (1997) SCAPE. Whereas the preceding experiment focused on mental effort expended during encoding, and how this effort predicts eventual memory performance, our next experiment was an investigation into the mental effort expended during recognition as a function of effort expended during encoding. Specifically, we examined recognition memory for spoken words, comparing pupil diameters to the same stimuli across study and test. To preface, we observed that, during accurate recognition trials, pupil diameters increased, relative to average diameters during encoding, and that this difference was greatest for more difficult stimuli (i.e., low-frequency words).

As in our ORE experiment, we hypothesized that memory could be predicted or reflected in pupil dilations, and that pupil diameters would differ depending on stimulus difficulty. To investigate this, we used words as stimuli, so that we could more precisely control inherent stimulus qualities (e.g., syllables, length, word frequency, etc.). To examine the pupillary reflex in response to words, we first examined basic word frequency effects in an experiment using a modified delayed naming procedure (see Papesh & Goldinger, 2009).

Thirty native English speakers from Arizona State University, with normal or corrected-to-normal vision, participated in exchange for partial course credit. Stimuli consisted of 150 common/uncommon word pairs (i.e., high frequency, HF/low frequency, LF), which were matched for visual features (e.g., few/pew). Words were presented on-screen for 500 ms each (black font on gray background), followed by a variable delay period of 250 – 2000 ms before a response tone. In the majority of trials (240), the response tone indicated to participants that they were to speak the word. In a minority of trials, however, the response tone signaled that participants were to abandon the speech plan and say "blah" instead. This secondary task allowed us to examine the effects of word frequency when production was equated. In a previous experiment of this type, Goldinger et al. (1997) observed that participants were faster to read HF words aloud, relative to LF words. We did not observe this effect. Instead, we observed that, at all trial stages following word perception, participants' pupils were more dilated to LF words, relative to HF words. That is, pupil dilations were a more sensitive reflection of underlying cognitive effort, relative to the behavioral index (naming latency).

To extend this finding, and to examine the pupillary reflex during a recognition memory task, we recently compared memory for HF and LF *spoken* words, eliminating the visual input entirely. The participants (N = 20) were all native speakers of English with no known hearing disabilities[3], all of whom participated in exchange for partial course credit. After being familiarized with the task, participants were asked to focus their gaze anywhere on the 17-inch Tobii 1750 monitor (with a solid background of gray, at a constant RGB of 150), so that continuous measurements could be taken of their pupil diameters. Before the study task, baseline diameter estimates were obtained for every

[3] This was inferred from accurate performance on a tone identification task prior to the experiment.

participant by instructing them to passively listen to a series of 12 words and sentences. The program automatically calculated a "baseline range" for each person, defined as any diameter falling within 2.5 standard deviations of their baseline stage diameter. During the study and test trials that followed, participants' pupil diameters were required to return to this range within 6 seconds, or the succeeding trial was dropped from analysis.

During study trials, participants passively listened to a series of 32 HF and LF words, with the timing of trials controlled by the speed with which each person's pupil diameters returned to the baseline range. Following a 3-minute distraction task (playing a computer game), participants listened to 64 spoken words, half of which were old, and indicated whether each word was "old" or "new." The voice of the speaker was constant across study and test.

Pupil dilations were trimmed for blinks and sorted into "trial events" corresponding to the fixation cross, the spoken word, and the blank screen (wherein the participants' pupils returned to the baseline range). We analyzed the data in a 2 (Word Frequency: HF/LF) x 2 (Presentation: first/second) x 2 (Accuracy: hit/miss) x 3 (Trial Event: fixation/word/blank) RM ANOVA. Although the Accuracy x Presentation interaction was not statistically significant ($F < 1$), a priori pairwise comparisons indicated that, during accurate recognition trials, pupils dilated to a greater extent during the second presentation (4.39 mm), relative to the first (4.24 mm), $F(1, 16) = 4.85$, $\eta_p^2 = .23$. Although this trend was evident for both HF and LF words, further pairwise comparisons indicated that this effect was driven primarily by LF words. When participants correctly recognized LF words, their pupils dilated to a greater extent during the second presentation (4.38 mm), relative to the first (4.23 mm), $F(1, 16) = 6.02$, $\eta_p^2 = .27$. Thus, as in our previous experiments, pupil dilation revealed underlying memory processing.

Conclusion: Production and Evaluation as "Stream of Consciousness"

Across all our studies, the heuristic processes predicted by SCAPE have been observed. Extending prior studies, our new results suggest that people engage in production and evaluation continuously, rather than engaging such processes only during a memory test. Moreover, our results show that these processes are accurately indexed by psychophysiological measures. The pupillary reflex reflects effort during both learning and recognition, which has promise for differentiating among theories of memory and assigns a concrete, dependent measure to the hypothesized stages of SCAPE. We believe the pupil reflex is sensitive enough to reflect fine gradations in the quality of memory and can indicate the use, or non-use, of various heuristic processes.

References

Bahrick, H. P., Bahrick, P. O., & Wittlinger, R. P. (1975). Fifty years of memory for names and faces: A cross-sectional approach. *Journal of Experiment Psychology: General, 104*, 54-75.

Beatty, J. (1982). Task-evoked pupillary responses, processing load, and the structure of processing resources. *Psychological Bulletin, 91*, 276-292.

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

Ekman, P., & Matsumoto, D. (1993). *Japanese and Caucasian Neutral Faces (JACNeuF)*. Photographs on CD-Rom.

Farah, M.J., Wilson, K. D., Drain, M. & Tanaka, J. N. (1998). What is "special" about face perception? *Psychological Review, 105*, 482-498.

Feenan, K. & Snodgrass, J. G. (1990). The effect of context on discrimination and bias in recognition memory for pictures and words. *Memory & Cognition, 18*, 515-527.

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.

Goldinger, S. D. , & Hansen, W. A. (2005). Recognition by the seat of your pants. *Psychological Science, 16*, 525-529.

Goldinger, S. D., He, Y., & Papesh, 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.

Haxby, J. V., Hoffman, E. A., & Gobbini, M. I. (2000). The distributed human neural system for face perception. *Trends in Cognitive Science, 4*, 223-233.

Henderson, J. M., Williams, C. C., & Falk, R. J. (2005). Eye movements are functional during face learning. *Memory & Cognition, 33,* 98-106.

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

Jacoby, L., Kelley, C. M. & Dwyan, J. (1989). Memory attributions. In H. Roediger III & F.I. M. Craik (Eds.), *Varieties of memory and consciousness: Essays in honour of Endel Tulving* (pp. 391-422). Hillsdale, NJ: Erlbaum.

Jacoby, L., & Whitehouse, K. (1989). An illusion of memory: False recognition influenced by unconscious perception. *Journal of Experimental Psychology: General*, *118*, 126-135.

Kleider, H. M., & Goldinger, S. D. (2004). Illusions of face memory: Clarity breeds familiarity. *Journal of Memory and Language, 50*, 196-211.

Kleider, H. M., & Goldinger, S. D. (2006). The generation and resemblance heuristics in face recognition: Cooperation and competition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 32*, 259-276.

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.

Leboe, J. & Whittlesea, B.W.A. (2002). The inferential basis of familiarity and recall: Evidence for a common underlying process. *Journal of Memory and Language*, *46*, 804-829.

Levin, D. T. (1996). Classifying faces by race: The structure of face categories. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*, 1364-1382.

Levin, D. T. (2000). Race as a visual feature: Using visual search and perceptual discrimination tasks to understand face categories and the cross-race recognition deficit. *Journal of Experimental Psychology: General, 129*, 559-574.

MacLin, O. H., & Malpass, R. S. (2001). Racial categorization of faces: The ambiguous race face effect. *Psychology, Public Policy, & Law, 7*, 98-118.

Meissner, C. A. & Brigham, J. C. (2001). Thirty years of investigating the own-race bias in memory for faces: A meta-analytic review. *Psychology, Public Policy, and Law, 7*, 3-35.

Neisser, U. (1967). *Cognitive psychology*. New York: Appleton-Century-Crofts.

Ng., W. & Lindsay, R. C. L. (1994). Cross-race facial recognition: Failure of the contact hypothesis. *Journal of Cross-Cultural Psychology, 25*, 217-232.

Papesh, M. H. & Goldinger, S. D. (2009). Pupil-Blah-Metry: Word frequency reflected in physiology. *Paper presented at the Western Psychology Association conference in Portland, OR (April, 2009)*.

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.

Porter, G. & Troscianko, T. (2003). Pupillary response to grating stimuli. *Perception, 32*, 156.

Roediger, H.L. III, & McDermott, K.B. (1995). Creating false memories: Remembering words not presented in lists. *Journal of Experimental Psychology: Learning, Memory and Cognition, 20*, 803-814.

Sporer, S. L. (2001). Recognizing faces of other ethnic groups: An integration of theories. *Psychology, Public Policy, & Law, 7*, 36-97.

Whittlesea, B.W.A. (1997). Production, evaluation and preservation of experiences: Constructive processing in remembering and performance tasks. In D. Medin (Ed.), *The Psychology of Learning and Motivation, 37* (pp. 221-264). New York: Academic Press.

Whittlesea, B.W.A., & Leboe, J. (2000). The heuristic basis of remembering and classification: Fluency, generation and resemblance. *Journal of Experimental Psychology: General, 129*, 84-106.

Whittlesea, B.W.A. & Williams, L.D. (1998). Why do strangers feel familiar, but friends don't? A discrepancy-attribution account of feelings of familiarity. *ACTA Psychologica, 98*, 141-165.

Whittlesea, B.W.A. & Williams, L.D. (2001). The discrepancy-attribution hypothesis: I. The heuristic basis of feelings of familiarity. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27*, 3-13.

Figure 1. Pupil dilation across learning trials. Separate functions are shown for learning of Asian and Caucasian faces, and for participants with relatively high or low memory scores.