

## Effects of visual and verbal presentation on cognitive load in vigilance, memory, and arithmetic tasks

JEFF KLINGNER,<sup>a</sup> BARBARA TVERSKY,<sup>b,c</sup> AND PAT HANRAHAN<sup>a</sup>

<sup>a</sup>Department of Computer Science, Stanford University, Stanford, California, USA

<sup>b</sup>Teachers College, Columbia University, New York, New York, USA

<sup>c</sup>Department of Psychology, Stanford University, Stanford, California, USA

### Abstract

Degree of pupil dilation has been shown to be a valid and reliable measure of cognitive load, but the effect of aural versus visual task presentation on pupil dilation is unknown. To evaluate effects of presentation mode, pupil dilation was measured in three tasks spanning a range of cognitive activities: mental multiplication, digit sequence recall, and vigilance. Stimuli were presented both aurally and visually, controlling for all known visual influences on pupil diameter. The patterns of dilation were similar for both aural and visual presentation for all three tasks, but the magnitudes of pupil response were greater for aural presentation. Accuracy was higher for visual presentation for mental arithmetic and digit recall. The findings can be accounted for in terms of dual codes in working memory and suggest that cognitive load is lower for visual than for aural presentation.

**Descriptors:** Pupil dilations, Cognition, Normal volunteers, Learning/memory

Assessing the cognitive load imposed by various visual tasks is important to the design of cognitively efficient visual interfaces. Most interfaces are visual, and many require shifting attention between a variety of tasks with varying loads on perception, attention, memory, and information processing. The psychophysiological study of cognitive load in this context requires a physiological proxy that responds to load quickly and reliably reflects small differences in load. One such proxy is the tendency of pupils to dilate slightly in response to cognitive loads (Loe-wenfeld, 1999).

This responsiveness of the pupil can provide detailed information about the timing and magnitude of cognitive loads and has thus been used to study a broad set of cognitive phenomena, including perception, memory, reasoning, and attention. (For general reviews, see Andreassi, 2006, ch. 12; Beatty, 1982b; Beatty & Lucero-Wagoner, 2000; Goldwater, 1972.)

This work was funded by the Stanford Regional Visual Analytics Center, through the U.S. Department of Energy's Pacific Northwest National Laboratory. Portions of this research were supported by NSF grants HHC 0905417, IIS-0725223, IIS-0855995, and REC 0440103. Our eye tracker was funded by the Stanford MediaX project and the Stanford School of Engineering. Some of the data from the aural presentation condition of all three experiments were presented at the Eye Tracking Research and Applications conference (Klingner, Kumar, & Hanrahan, 2008).

The experiments described in this article were conducted in compliance with the policies of Stanford's institutional review board. All participants gave informed consent, and their rights and safety were protected.

Address correspondence to: Jeff Klingner, Gates Building Room 3B-396, Stanford, CA 94305-9035, USA. E-mail: klingner@stanfordalumni.org

Kahneman (1973) used pupillary dilations as the primary empirical foundation for his attention theory of effort. He identified three criteria desirable in general for physiological proxies for effort and that he observed in pupillary dilations: differences in the magnitude of averaged pupillary dilations reliably reflect (a) different difficulty levels of a single task, (b) differences in difficulty across qualitatively different tasks, and (c) individual differences in ability. In a review nine years later, Beatty (1982b) reaffirmed that the experimental evidence then available did indeed show that pupillary dilations fulfill all three of Kahneman's criteria.

Nearly all the studies that had been done at that time used auditory stimuli, in order to avoid interference from the pupillary light reflex. Since then, several investigators have successfully controlled for reflex dilations while using visual stimuli and shown that task-induced dilations can serve as reliable proxies for cognitive load in visual tasks such as reading (Just & Carpenter, 1993) and visual search (Bucks & Walrath, 1992; Porter, Troscianko, & Gilchrist, 2007). These studies all show that the magnitudes of momentary pupil dilations reliably reflect different levels of difficulty within individual tasks and have thus validated pupillary dilations as continuing to fulfill Kahneman's first criterion even on extension to visual tasks.

However, to our knowledge, nobody has examined the effect of aural versus visual presentation mode itself on the magnitude of pupillary dilations. This lack of data confounds the use of dilations for comparing cognitive loads between visual and aural tasks, because it cannot be known how much of the difference is caused by the difference in presentation modalities and how much is caused by differences in postperception task demands. In other words, it is still not known whether Kahneman's second

criterion, intertask comparability, is still fulfilled by pupil dilations when they are used to study visual as well as aural tasks.

Developments in graphics have brought interfaces, newspapers, textbooks, and instructions that increasingly present changing visual information. Viewers need to attend to, search through, and evaluate this information in order to integrate it. Are visual interfaces the best way to present this information, or might cognitive load be lessened with auditory presentation? Are the parameters of cognitive load similar for visual and auditory presentation? To address those questions, the present experiments compare visual and auditory presentation for three classic paradigms in attention and perception.

In choosing tasks, we sought to (a) span diverse types of cognitive load, (b) replicate well-studied tasks to enable comparisons to prior results, and (c) use simple stimuli that are easy to match between aural and visual presentation. We chose mental multiplication (Experiment 1), digit-span memory (Experiment 2), and vigilance (Experiment 3).

In our experimental designs, we took care to control for all known noncognitive pupillary reflexes. The aural and visual conditions employ visual fields with matching brightness and contrast and with stimulus onset effects controlled; the difference is that in the aural conditions, the task-relevant stimuli were heard, and in the visual conditions they were seen.

Because visual perception is generally believed to involve less effort, but the subsequent central processing demands were matched between the two presentation conditions, we expected dilations evoked by visually presented tasks to start out smaller but to eventually reach the same peak dilation as those evoked by the aurally presented versions. We also expected this difference in effort to be reflected in lower error rates and quicker responses in the visual conditions.

## EXPERIMENT 1: MENTAL MULTIPLICATION

Hess and Polt (1964) triggered broad interest in cognitive pupillometry when they reported that solving mental multiplication problems caused pupil dilations and that harder problems evoked larger dilations. Their results were replicated by Bradshaw (1968) for mental division with remainders; Boersma, Wilton, Barham, and Muir (1970) for mental addition in a study of mental retardation; and Ahern and Beatty (1979) in a study of the effect of individual differences in ability as measured by SAT scores. Recently, Marshall (2002) used a mental arithmetic task to validate a wavelet-based method of analyzing pupil measurements.

These studies variously investigated the effects of problem difficulty, response mode, and participant ability on pupil dilations, but none investigated the influence of aural versus visual stimulus presentation. Experiment 1 was a replication of Ahern and Beatty's (1979) mental multiplication study with the addition of two visual stimulus presentation conditions, one with timing matched to the aural condition, and one with simultaneous visual presentation of both parts of the multiplication problem.

### Method

Unless otherwise specified, method details described here apply to all three experiments.

### Participants

Twenty-four Stanford undergraduates participated in this experiment. All had normal or corrected-to-normal vision. We excluded participants with contact lenses or eyeglasses providing an astigmatism correction or a refractive correction greater than 10 diopters, which can interfere with accurate pupil diameter tracking. We compensated participants with Amazon.com gift certificates. The value of each participant's gift certificate depended on his or her task performance and varied from about \$15 for the lowest scores to about \$35 for the highest. Such monetary incentive was shown by Heitz, Schrock, Payne, and Engle (2008) to increase the magnitude of pupillary responses.

### Apparatus

We measured the size of participants' pupils using a Tobii 1750 remote eye tracker (Tobii Technologies, 2007). This device is designed primarily to track people's gaze direction, but its method of gaze tracking also enables high-speed pupillometry (Klingner, Kumar, & Hanrahan, 2008). The eye tracker is based on a standard LCD computer display, with infrared lights and a high-resolution infrared camera mounted at the edges of the screen. This remote-camera setup requires neither a chin rest nor a head-mounted camera, enabling pupil measurements without encumbrance or distraction. Measurements are corrected for changes in apparent pupil size due to head motion toward or away from the camera. Accurate pupil tracking with this equipment requires a head motion speed of less than 10 cm/s within a head box of about  $30 \times 15 \times 20$  cm at our initial seating distance of 60 cm from the screen.

We placed the eye tracker on a desk with the top of the screen approximately 140 cm from the floor. Participants sat in a chair adjusted so that their eyes were at this same height. Participants initiated trials and gave task responses using a two-button computer mouse on the desk between them and the eye tracker. We used a relatively bright room, with  $27 \text{ cd/m}^2$  of luminance from the surrounding walls at eye level and 32 lx incident at participants' eyes.

### Data Processing

Under infrared illumination, participants' pupils appear as bright ovals in the eye tracker's camera image. The Tobii 1750 measures the size of a participant's pupil by fitting an ellipse to the pupil image then converting the width of the major axis of that ellipse from pixels to millimeters based on the measured distance from the camera to the pupil. Due to inaccuracy in this measurement of camera-pupil distance, measurements of absolute pupil size may have errors of up to 5%, but sample-to-sample changes in pupil diameter are much more accurate. This better accuracy for relative measures makes eye trackers well suited for cognitive pupillometry, where the measurement of interest is usually changes in pupil diameter relative to their diameter at the end of an accommodation period preceding each trial (Beatty & Lucero-Wagoner, 2000). This measure has been found to be independent of baseline pupil diameter and commensurate across multiple laboratories and experimental procedures (Beatty, 1982b; Bradshaw, 1969, 1970).

Our apparatus samples pupil size at 50 Hz, with each sample measuring both eyes simultaneously. Because the left and right eyes exhibit matching pupillary responses, we used the average of the two eyes' pupil diameters to reduce measurement noise. During moments when an eyelid, eyelash, or eyeglasses frame blocked the camera's view of one pupil, we used the other pupil

alone. We performed standard baseline subtraction in each trial based on the average pupil diameter measured over 20 samples (400 ms) at the end of a prestimulus accommodation period. After filling blinks via linear interpolation, we smoothed the raw pupil signals with a 10-Hz low-pass digital filter. We constructed the pupil traces shown in all figures by stimulus aligning and averaging all trials for each illustrated condition.

*Data processing for statistical evaluation of differences in dilation magnitude.* We quantified dilation magnitudes with the mean amplitude method (Beatty & Lucero-Wagoner, 2000, p. 148; Handy, 2004, p. 38). This method involves first measuring a baseline pupil size for each trial by averaging pupil size during a prestimulus accommodation period, then computing the average pupil size relative to this baseline during a response window defined for each task. We chose the mean dilation quantification method over the also common peak dilation method, because the latter is more sensitive to noise. We quantified each trial separately, enabling statistical evaluation of effect size and significance.

*Significance tests.* We used an alpha level of .05 for all statistical tests. Tests of differences in mean dilation magnitude were all based on partitions of variance (ANOVA). Following the policy of Jennings (1987), we applied the Huynh and Feldt (1980) correction to degrees of freedom for within-subjects factors with more than two levels. In such cases, we report the Huynh–Feldt nonsphericity correction parameter  $\hat{\epsilon}$ , the uncorrected degrees of freedom, and the corrected  $p$  value. We evaluated the significance of differences in error rates through one-tailed tests for equality of proportions with Yates' continuity correction (Miettinen & Nurminen, 1985).

*Data publication.* All the raw data we collected, including practice trials and all excluded data, as well as the source code for the computer programs we used to collect and analyze it, are published as supporting information in the online version of this article. (See details at end of article.)

### **Controlling for Noncognitive Pupillary Motions**

*Pupillary light reflex.* The largest potentially confounding pupillary motion is the pupillary light reflex, which is much larger in magnitude than cognition-induced pupil changes (Loewenfeld, 1999). We followed standard practice (e.g., Moresi et al., 2008; Verney, Granholm, & Dionisio, 2001) in maintaining constant visual field luminance across experimental conditions. Additionally, we used prestimulus masks equal in luminance and contrast to the stimulus, to avoid luminance and contrast changes at stimulus onset.

*Luminance changes caused by shifting gaze.* Experiments in which participants shift their gaze to look at many parts of a visual stimulus, including studies of reactions to photographs (Dabbs & Milun, 1999; Libby, Lacey, & Lacey, 1973), visual search (Bucks & Walrath, 1992; Porter et al., 2007), and visual scanning (Pomplun & Sunkara, 2003; Van Orden, Limbert, Makeig, & Jung, 2001) are subject to pupillary light reflexes when participants fixate on local areas of the stimulus with varying luminance even though the overall luminance of the stimulus does not change. Reading studies, in which textual stimuli have relatively uniform local luminance and consistent fixation sequences, are not as vulnerable to this problem and have successfully measured small task-evoked pupillary responses amidst

active eye movements (e.g., Just & Carpenter, 1993). We controlled for saccade-induced luminance changes by presenting all stimuli at a fixed location within an area small enough to fall within the fovea and by helping participants to keep their gaze fixed by presenting a fixation target at all times and keeping trial durations under 20 s.

*Other visual causes of pupil changes.* In addition to the light reflex and the cognitive load response, the pupil also exhibits small dilations or contractions in response to changes in accommodation distance (Loewy, 1990), contrast (Ukai, 1985), spatial structure (Cocker, 1996) and the onset of coherent motion (Sahraie & Barbur, 1997). Kohn and Clynes (1969) showed that simply changing the color content of a visual stimulus, without changing either local or global luminance, can cause the pupils to either dilate or contract, depending on the nature of the color change. We controlled for all of these influences on pupil size by using achromatic, fixed-distance, nonmoving, constant-contrast stimuli.

*Pupillary blink response correction.* We followed the common practice of filling gaps in the data caused by blinks with linear interpolation. However, by performing blink-locked averaging of data from a pilot study, we observed that blinks result in changes in pupil diameter that persist for a few seconds after the blink. This pupillary blink response consists of a very brief dilation of about 0.04 mm, followed by a contraction of about 0.1 mm and then a gradual recovery to preblink diameter over the next 2 s. The timing and magnitude of these changes depend on the duration of the blink. To the extent that blinks occur randomly, pupillary blink responses add noise to averaged pupil diameter measurements, and to the extent that blinks are correlated with stimuli, pupillary blink responses add bias to averaged pupil diameter measurements.

We gathered data from 20 thousand binocular blinks that occurred during several of our eye-tracking studies, grouped the blinks by duration, and averaged them to determine 3-s-long blink response correction signals. We then removed the pupillary blink responses in the current study by altering the data following each blink by subtracting the blink response correction signal corresponding to the length of that blink.

For stimulus-correlated blinks, the general effect of this correction is to decrease the magnitude of pupillary responses measured in the first second following a blink by about 0.03 mm and increase the magnitude of pupillary responses measured in the second second following a blink by about 0.05 mm. For stimulus-uncorrelated blinks, the general effect of this correction is to remove measurement noise and thereby decrease the standard errors of the mean in stimulus-locked averages of dilation magnitude.

Because this is a new data processing technique for pupil data, we reran the analyses presented here without blink response correction and found that the correction did not change the significance of any of our results and changed the effect sizes by only 0.005–0.01 mm, suggesting that blinks were not well correlated with stimuli for the tasks we examined and contributed only noise to the stimulus-locked averages.

### **Stimuli**

Stimuli for all experiments were numbers between 1 and 20. Under the aural condition, stimuli were 500-ms digitized recordings of spoken numbers played over a computer speaker placed

directly behind the screen. Under the visual condition, we displayed these numbers at the center of the eye tracker's integrated 17-in.  $1280 \times 1024$  LCD screen. We used a 28-point font size so that the digits spanned  $0.73^\circ$  (about a third of the foveal span) when viewed from participants' initial seating distance of 60 cm. These numerals were black, and the rest of the screen was always filled with a uniform background of  $64 \text{ cd/m}^2$  medium gray.

The onset timing and duration of stimuli presentation under the aural and visual conditions was matched. During periods of time with no stimulus (between trials, during the prestimulus pupil accommodation period, and in between presentation of numbers during the task), we presented silence in the aural condition and masked the stimulus by displaying an "X" at the center of the screen in place of a number, in order to remove contrast and brightness changes caused by the appearance or disappearance of the numerals. The absence of clear constrictions following the time of visual stimulus change in the visual waveforms provides evidence that these stimulus changes per se had little effect on the pupil in our experiments.

### Procedure

Before each task, we explained the task to participants, then allowed them to practice until they were familiar and comfortable with the task presentation and providing their responses.

All trials were initiated by participants, who first fixated a small target at the center of the screen before starting the trial by clicking a mouse button. Participants' gaze thus remained at the center of the screen for the duration of each trial and during most of the short intervals between trials. A run of trials for a single task generally took about 5 min. We told participants that they could take breaks at any point between trials to rest their eyes; 2 did so.

In Experiments 1 and 2, where the tasks required numerical responses, we asked participants to type their responses into a low-contrast on-screen keypad. We did this to automate data collection and to avoid pupillary reflexes to varying brightness caused by looking away from the screen. Because button-press responses themselves induce pupillary responses (Richer & Beatty, 1985), and we could not avoid such interference by using spoken responses (Bradshaw, 1967; Kahneman, Onuska, & Wolman, 1968), we limited our analysis to pre-response periods.

### Task Description

We began each trial with a 2-s prestimulus accommodation period, during which participants rested their eyes on a fixation target in the center of the screen in order to stabilize their pupils. We then presented the participant with two numbers, the multiplicand and multiplier, separated by 2 s. Five seconds after we presented the multiplier, we asked participants for the two numbers' product. In a departure from Ahern and Beatty's (1979) procedure, rather than speaking the product, participants typed it into an on-screen keypad using the mouse. For significance testing, we averaged pupil dilation across the 5-s window between multiplier presentation and participant response.

For each trial we randomly selected a difficulty level of easy, medium, or hard, then chose the multiplier and multiplicand randomly according to Ahern and Beatty's (1979) definition of these difficulty levels: easy problems took the form  $\{6, 7, 8, 9\} \times \{12, 13, 14\}$  (e.g.,  $7 \times 13$ ), medium were  $\{6, 7, 8, 9\} \times \{16, 17, 18, 19\}$ , and hard  $\{11, 12, 13, 14\} \times \{16, 17, 18, 19\}$ . Nine participants had memorized the multiplication table through  $12 \times 12$

and the rest through  $10 \times 10$ , so all but a few of the easiest problems required mental computation beyond simple recall.

In the visual presentation condition, we used two timing variants. In the sequential treatment, which replicates Ahern and Beatty (1979), the multiplicand and multiplier were presented one after the other with timing matched to the aural condition. In the simultaneous treatment, both numbers were shown on the screen together for the full 8 s between the prestimulus accommodation period and the response prompt. This simultaneous and continuous presentation was intended to remove the requirement that subjects quickly read and remember the short-lived stimuli and thereby isolate the cognitive load imposed by mental multiplication from that caused by remembering the numbers.

We instructed participants not to provide a response in cases when they forgot one of the two numbers or gave up on computing their product. This occurred in 10% (65/632) of the trials, mostly for hard problems. Because these trials did not involve mental multiplication, we excluded them from analysis.

### Results

Our results for aurally presented mental multiplication problems matched those of Ahern and Beatty (1979). We observed two peaks in pupil dilation: a brief, small dilation following presentation of the multiplicand and a longer, larger dilation following presentation of the multiplier, during the time when participants computed the numbers' product.

#### Dilation Magnitude by Presentation Mode

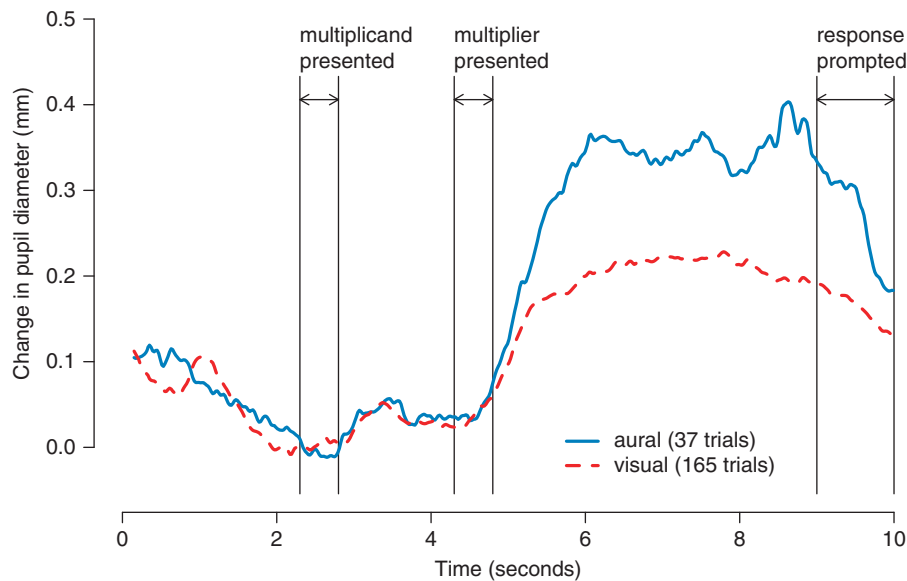
Presentation mode affected the overall magnitude of pupil dilations but not their qualitative shape. The onset timing, duration, and overall shape of pupil dilations caused by mental multiplication was the same for both auditory and visual presentation. The size of participants' dilations, however, was significantly larger in the auditory condition ( $M = 0.35 \text{ mm}$ ,  $SD = 0.11 \text{ mm}$  vs.  $M = 0.16 \text{ mm}$ ,  $SD = 0.13 \text{ mm}$ ),  $F(1,22) = 12.1$ ,  $p = .002$ . This difference in magnitude is clear in Figure 1, which shows the pupil dilation evoked by the mental multiplication task, averaged across all trials and participants and broken down by task presentation mode.

#### Dilation Magnitude by Task Difficulty

Consistent with prior investigations of mental arithmetic, we found a clear difficulty effect on dilation magnitude. Easy multiplication problems caused the smallest pupil dilations ( $M = 0.17 \text{ mm}$ ,  $SD = 0.19 \text{ mm}$ ), hard problems the largest ( $M = 0.27 \text{ mm}$ ,  $SD = 0.16 \text{ mm}$ ), with dilations to medium problems in between ( $M = 0.21 \text{ mm}$ ,  $SD = 0.15 \text{ mm}$ ). These differences were significant,  $F(2,30) = 13.1$ ,  $p = .0008$ ,  $\bar{\epsilon} = .67$ .

#### Pupillary Response to Continuously Visible Problem

The pupil dilation evoked by problems with both components visible simultaneously for 8 s had a different pattern: a single long dilation and contraction, rather than the two peaks we observed in the sequential case. In addition, the mean pupil dilation was smaller in the simultaneous case ( $M = 0.13 \text{ mm}$ ,  $SD = 0.11 \text{ mm}$  vs.  $M = 0.30 \text{ mm}$ ,  $SD = 0.13 \text{ mm}$ ),  $F(1,22) = 10.3$ ,  $p = .004$ . This result is not surprising, because the simultaneous-presentation trials lack a second stimulus event to cause a second peak, and these trials were easier to solve, because they did not require participants to remember the two presented numbers.



**Figure 1.** Average pupil dilation evoked by visually and aurally presented mental multiplication problems. The two presentation modes elicited dilations with similar timing, duration, and shape, but different magnitudes. Vertical lines show the times during which the two numbers were spoken or displayed and the time during which the participants responded.

#### Task Performance by Presentation Mode

Participants made significantly more errors on aurally presented problems (40%) than visually presented problems (25%);  $\chi^2(1, N = 632) = 3.39, p = .03$ .

#### Discussion

This experiment compared cognitive load under aural and visual presentation of mental arithmetic problems. The overall pattern of task-evoked pupil dilations was similar in both conditions and replicated previous aural work. Intriguingly, both the better performance under visual presentation and greater cognitive load under aural presentation suggest an advantage for visual presentation of mental arithmetic. This may be because poststimulus visual persistence alleviates some load on working memory.

### EXPERIMENT 2: DIGIT SPAN SEQUENCE MEMORY

Short-term recall of a paced sequence of digits (also known as the digit span task) is the most popular experimental task in cognitive pupillometry. First reported by Kahneman and Beatty (1966), the task was also used to investigate the related processes of long-term recall (Beatty & Kahneman, 1966), grouping (Kahneman, Onuska, & Wolman, 1968), and rehearsal (Kahneman & Wright, 1971). Peavler (1974) showed that the pupil reaches a plateau dilation of about 0.5 mm around the presentation of the seventh digit. Granholm, Asarnow, Sarkin, and Dykes (1996) replicated this finding, confirming that pupil dilation averaging can be used to estimate both the momentary load and the maximum capacity of working memory.

As with the mental multiplication task, all prior investigations of pupil dilations evoked by this task presented the digit sequence aurally. Our experiment is a replication of the original Kahneman and Beatty (1966) study, with the addition of a visual presentation condition.

#### Method

Details regarding study participants, equipment, and procedures not specific to this task are described in the Method section within the description of Experiment 1.

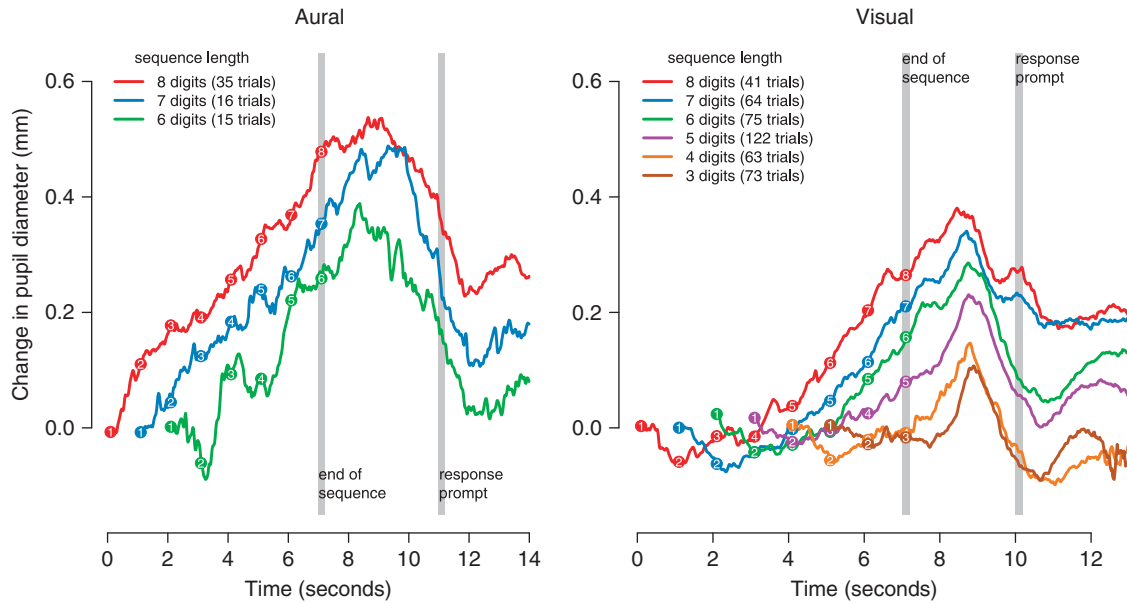
As with Experiment 1, we started each trial with a pupil stabilization period in which we measured baseline pupil diameter. We then presented a sequence of digits at the rate of one per second, either spoken aloud or displayed on the screen. After a brief retention pause, participants then reported back the sequence using an on-screen keyboard as in Experiment 1. We used the first 2 s of this retention pause as the response window for pupil diameter averaging and significance testing, because this is the moment when Kahneman and Beatty (1966) observed maximum dilations. We randomly varied the length of the presented sequence for each trial independently between six and eight digits for aural presentation and between three and eight digits for visual.

#### Results

Averaged pupil traces from this experiment are shown in Figure 2. Under both aural and visual presentation, changes in pupil diameter followed the same qualitative pattern observed by Kahneman and Beatty's (1966) aural study: Participants' pupils gradually dilated as the digits were memorized, reached a peak 2 s after the final digit during the pause while the sequence was retained in memory, then gradually contracted as the participants reported the digits back.

#### Dilation Magnitude by Presentation Mode

Aural presentation caused significantly larger pupil dilations during the retention pause than visual presentation ( $M = 0.44$  mm,  $SD = 0.22$  mm vs.  $M = 0.24$  mm,  $SD = 0.17$  mm),  $F(1,20) = 5.9, p = .02$ .



**Figure 2.** Pupil dilation evoked by a digit-span memory task presented aurally (left) and visually (right). The two charts are aligned and plotted at the same vertical scale. The numbered circles on each line show the times at which each digit was spoken (aural presentation) or displayed (visual presentation). The curves are each shifted horizontally so they are aligned at the end of the stimulus sequence. Thus the longest sequence (eight digits) starts the furthest to the left. Aural presentation caused larger dilations than visual, and under both presentation modes, longer memorization sequences elicited larger pupil dilations.

#### *Dilation Magnitude by Task Difficulty*

We found a significant effect of sequence length on the magnitude of pupil dilations during the retention pause,  $F(3,60) = 3.73$ ,  $p = .02$ ,  $\bar{\epsilon} = .96$  (see Figure 2). The magnitude of the dilation increased monotonically with the length of the memorized sequence.

#### *Task Performance*

Considering sequences of all lengths, participants made significantly more recall errors under aural (30%) than visual (24%) presentation,  $\chi^2(1, N = 1232) = 3.94$ ,  $p = .02$ , though this result is reversed if only the longest (length 7 and 8) sequences are considered. Average digit span was 6.0 digits for aural presentation and 5.6 digits for visual.

#### **Discussion**

This experiment compared cognitive load under short-term memorization of aurally and visually presented digit sequences. As with the mental arithmetic task, the qualitative shape of average pupil dilations was similar in both presentation modes, but the magnitude of dilations was smaller under visual presentation.

Although visual presentation led to significantly greater overall performance, the difference was not large, and rates of recall for the longer sequences and average digit span scores suggest a small performance advantage for aural presentation. A general advantage to serial recall under aural presentation, especially for items late in the sequence, is well documented (Greene, 1992, p. 22; Penney, 1989; but see Beaman, 2002). Our findings on recall performance are mixed, but the larger dilations we observed in the aural condition suggest that task performance in this mode comes with the cost of higher cognitive load.

#### **EXPERIMENT 3: VIGILANCE**

The mental multiplication and digit span tasks are both strongly dependent on working memory. We designed our third experiment to investigate the effect of aural versus visual stimulus mode on pupil dilations evoked by less memory-dependent processes, using a task that requires intermittent vigilance, stimulus discrimination, and speeded motor responses.

#### **Method**

Details regarding study participants, equipment, and procedures not specific to this task are described in the Method section within the description of Experiment 1.

In each trial, we presented an ascending sequence of numbers from 1 through 20. We told participants that the sequence might progress normally or might contain errors at the number 6, 12, and/or 18. When they noticed an error (a target), they were to push a button as quickly as possible. For example, part of the sequence might be "... 10, 11, 12, 13, ...," in which case we instructed the participants to do nothing, or it might be "... 10, 11, 7, 13, ...," in which case we told them to push the button as soon as possible after noticing the "7." We inserted sequence errors (targets) at the three possible positions independently and randomly with probability one half. Thus, any trial could contain zero, one, two, or three targets, and participants knew exactly when the targets might appear.

In the aural condition, "6" was never replaced by "16," nor "18" by "8," so that errors were apparent from the start of each spoken target stimulus. As with Experiments 1 and 2, the aural and visual conditions were matched on visual field luminance, contrast, and stimulus timing.

Unlike Experiments 1 and 2, this experiment did not replicate a past study, though it incorporated aspects of prior experiments. Beatty (1982a) found pupil dilations evoked by target tones in an auditory vigilance task, though in that experiment target locations were randomized, so that participants could not anticipate them, and continuous rather than intermittent vigilance was required. The anticipated increase in vigilance required by this task was studied by Richer, Silverman, and Beatty (1983).

## Results

### *Dilation Magnitude by Presentation Mode*

Figure 3 shows the average dilation evoked by the vigilance task, comparing aurally and visually presented trials. Both conditions elicited strong dilation peaks beginning about 1 s before and peaking 500–1000 ms after each moment when participants were alert for mistakes in the counting sequence. The 1-s anticipatory dilation is consistent with measurements of the readiness potential made using scalp electrodes by Becker, Iwase, Jürgens, and Kornhuber (1976), who found evidence of motor preparation beginning a bit more than 1 s before action, and is shorter than the 1.5-s lead observed by Richer et al. (1983) before the presentation of an action-determining stimulus.

For significance testing, we used a wide response window, starting 3 s before each moment when a target could occur and ending 3 s after, encompassing both the prestimulus anticipatory dilation and the poststimulus motor-response peak. The mean dilation in the auditory presentation condition ( $M = 0.096$  mm,  $SD = 0.048$  mm) was significantly larger than for visual presentation ( $M = 0.057$  mm,  $SD = 0.046$  mm),  $F(1,23) = 7.93$ ,  $p = .01$ .

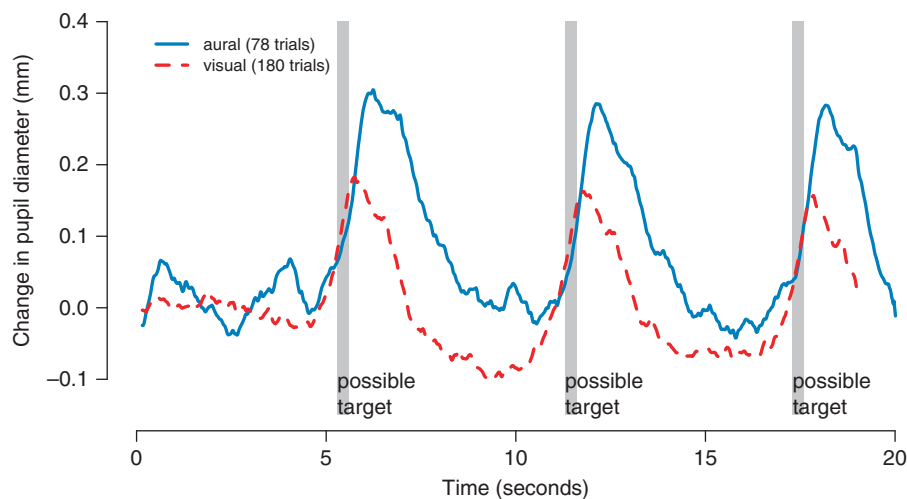
### *Dilation Onset and Peak Latency by Presentation Mode*

In contrast to Experiments 1 and 2, the three task repetitions in each of Experiment 3's trials effectively tripled the number of trials available for analysis and so provided enough data to pinpoint the peak dilation precisely in time and revealed a minor timing difference between the dilations for aural and visual

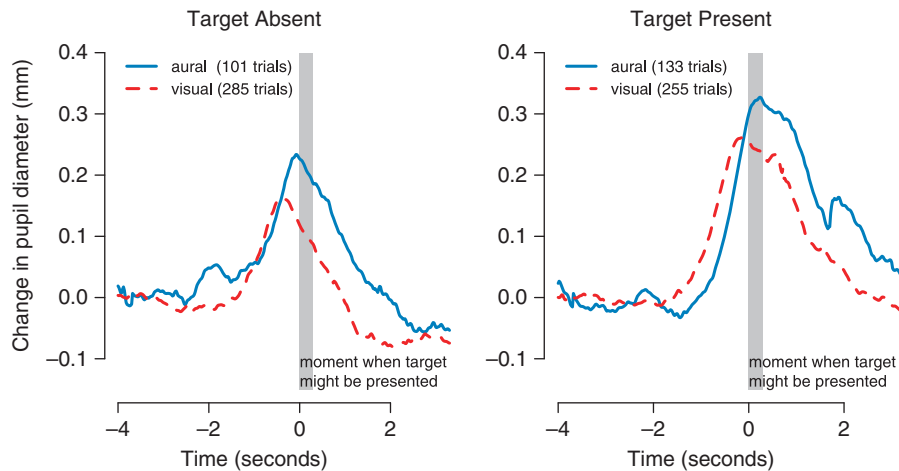
vigilance. Whether the target was present or absent, the dilation began and peaked slightly later under aural presentation (see Figure 4). This slightly later dilation evoked by aural stimulus was probably due to the time taken for the stimulus to be presented, because hearing is generally believed to have lower latency than vision (Misulis & Fakhoury, 2001; Welford, 1980). This interpretation is consistent with the difference in mean reaction time we observed: 410 ms ( $SD = 111$  ms) for visual presentation and 713 ms ( $SD = 140$  ms) for aural.

### *Dilation Magnitude and Timing by Target Presence*

At every potential mistake point, whether or not a target is present, this task required heightened vigilance, motor response preparation, and comparison of the presented number with the expected correct sequence number. We therefore expected dilations in both cases to be similar, perhaps with slightly larger or longer dilations in cases where targets actually appeared, caused by error recognition, the additional requirement of carrying out the motor response, or both. We checked this hypothesis by grouping all time segments surrounding moments when the targets were present and averaging them separately from those when the targets were absent. The resultant pupil dilation averages are shown in Figure 4. Pupil dilations evoked by targets were larger and longer than those measured during moments when targets were possible but did not appear ( $M = 0.10$  mm,  $SD = 0.046$  mm vs.  $M = 0.037$  mm,  $SD = 0.047$  mm),  $F(1,23) = 22.8$ ,  $p < .0001$ . The averaged pupil diameter trace for cases with a target (right side of Figure 4) showed a secondary peak about 1.5 s after the target appeared. Because mean response time was 515 ms ( $SD = 188$  ms), the latency between response and this secondary peak was about 1 s. Because Richer and Beatty (1985) observed similar dilation-response latencies in a nonreactive button pushing task and because this secondary peak was only present when motor response was required, we interpreted the secondary peak as an artifact of that motor response. The interaction of stimulus mode and target presence was not significant,  $F(1,23) = 0.351$ ,  $p = .6$ . The larger dilations evoked by aural task presentation persist whether a target is present or absent (see Figure 4).



**Figure 3.** Average pupil dilations evoked by a vigilance task presented aurally and visually. The vertical gray bars show the moments at which participants were vigilant for mistakes in a counting sequence (“targets”). The aurally presented task led to larger dilations, but the two presentation modes elicited dilation profiles with similar shape and timing.



**Figure 4.** Target effect on pupil dilations evoked by heightened vigilance. The data shown are the same as those in Figure 3. Each trial of Experiment 3 had three moments at which we told participants to expect possible targets, which occurred independently at each moment with probability one half. The chart on the left shows the mean dilation in moments in which a target did not occur, and the chart on the right shows the mean dilation in moments when a target did occur. Targets elicited longer and larger pupil dilations, with a secondary peak about 1.5 s after target presentation. This secondary peak corresponds to the motor activity of responding to the target's presence. Whether a target was present or absent, dilations were larger in the aural condition, and the peak dilation under aural presentation occurred about half a second later than with visual presentation.

### Task Performance

Participants made more errors in the counting vigilance task when it was presented aurally (8.5%) than visually (6.1%), but this difference was not significant,  $\chi^2(1, N = 774) = 7.80, p = .14$ .

### Discussion

This experiment compared the cognitive load under aurally and visually presented intermittent vigilance tasks. As with the other two tasks we studied, the two presentation modes elicited pupil dilations with very similar timing and overall shape, and although we did not observe a significant performance difference, visual presentation caused lower cognitive load.

In addition to the presentation mode effect, we also observed that the presence of targets was associated with larger pupil dilations. This difference is consistent with the additional cognitive demand of pushing the button in cases when the target is present.

## GENERAL DISCUSSION

### Summary of Experiments

In our first experiment, participants completed mental multiplication problems either spoken aloud or displayed on a computer screen. Our second experiment examined the digit span short-term recall task, again presented both aurally and visually, and our third experiment considered a speeded-reaction vigilance task that did not rely heavily on working memory. In all tasks, we controlled the stimulus timing between the two modes as well as controlling all aspects of the visual field—brightness, contrast, and participant fixation—in order to minimize noncognitive pupillary reactions.

### Summary of Findings

We found that the pupil dilations evoked by all three tasks were qualitatively similar under auditory and visual presentation, but that auditory presentation led to larger pupillary dilations.

### Qualitative Match

In all three of our experiments, we observed that pupil dilations in both modes had about the same onset timing, duration, and overall shape (see Figures 1, 2, and 3). Additionally, in the two tasks that replicated classic pupillary response studies, mental multiplication (Hess & Polt, 1964) and digit span (Kahneman & Beatty, 1966), we also found a qualitative match between the dilations we observed and the aural-only classic results. Both of these qualitative correspondences—visual to aural in our experiments and visual to classic aural findings—suggest that the pupil dilations we observed to visually presented tasks reflect the cognitive demands of the tasks and were generally free of distortion caused by noncognitive pupillary reactions to brightness or contrast changes.

### Quantitative Difference

In all three of our experiments, we observed significantly larger pupillary dilations when we presented tasks aurally than when we presented them visually. The differences were 0.19 mm (0.35 mm vs. 0.16 mm) for mental multiplication, 0.18 mm (0.43 mm vs. 0.25 mm) for digit span memory, and 0.08 mm (0.23 mm vs. 0.15 mm) for vigilance.

### Implications

Because we were careful to control for noncognitive pupillary responses caused by brightness, contrast, and so forth and because of our finding of a qualitative match in dilation trajectories between conditions, we believe that the difference in magnitude between the two conditions was a result of differences in cognitive load. We therefore interpret this result as evidence that visual task presentation leads to lower cognitive load than auditory presentation across all three of the tasks we studied.

This finding contradicted our hypothesis that similar task demands would lead to similar magnitude dilations in the two cases, perhaps with an initially smaller dilation under visual presentation caused by the lesser difficulty of seeing versus hearing numbers. Instead, we found that aural task presentation led to



larger pupil dilation not only during initial stimulus comprehension but also throughout task completion.

Taken together with the better performance we observed in the visual conditions, this finding indicates that visual presentation facilitates processing for all three tasks. That is, comprehending and remembering numbers is easier when they are seen than when they are heard.

#### **Relation to Prior Digit Span Findings**

In the case of digit span, our finding of an advantage for visual presentation seemed to contradict prior studies that found better performance under aural task presentation. Improved recall of heard numbers relative to seen numbers is very well established (Greene, 1992, p. 22; Penney, 1989; but see Beaman, 2002). Indeed, in our measurements of error rates, we found that although visual presentation led to significantly greater overall performance, the difference was not large, and rates of recall for the longer sequences and average digit span scores suggest a small performance advantage for aural presentation, as was found in the cited investigations.

This apparent contradiction between lower cognitive load under visual presentation and superior recall of heard numbers can perhaps be resolved by drawing a distinction between levels of effort and levels of performance (cf. Paas & Van Merriënboer, 1994). Although performance was better for heard numbers, our pupillary data suggest that this greater performance may have come with the cost of greater effort and cognitive load.

#### **Relation to Prior Mental Arithmetic Findings**

Prior investigations of mental arithmetic have not often addressed the effect of stimulus mode. In a study of the relative importance of different components of working memory in serial mental addition, Logie, Gilhooly, and Wynn (1994) observed that visual problem presentation led to better performance and less degradation in the context of a variety of interfering tasks. Our finding of better performance in the visual case matches theirs. They concluded that the central executive, the visuospatial store, and subvocal rehearsal are all involved in mental arithmetic. Taken together with these data, our finding of lower cognitive load in the visual case suggests that visual presentation facilitates mental arithmetic performance by aiding the recruitment of all three of these components of working memory. This possibility is supported by recent functional magnetic resonance imaging data collected by Fehr, Code, and Herrmann (2008), who found that presentation mode can significantly impact which regional

neuronal networks are employed in the calculation process for mental arithmetic.

#### **Conclusion**

It is well known that visual presentation can lead to higher performance on complicated tasks such as schema learning (Clark & Paivio, 1991) and finding patterns in data (Chen, 2004). Such advantages are typically attributed to the benefits of a persistent external representation that reduces load on working memory. Our finding of a visual advantage even for simple tasks and even though we controlled presentation duration, displaying the digits exactly as long as they took to speak, suggests that something besides visual persistence underlies this visual advantage.

One account for superior performance under visual rather than auditory presentation rests on the role of dual codes in working memory (e.g., Baddeley, 2007; Paivio, 1990). Visual presentation is likely to encourage dual coding of the stimuli (e.g., Paivio, 1990). Extensive research has shown that having two mental representations for something, notably, both visual and verbal, is better for memory than having one. If one internal representation is lost or corrupted, the other can compensate. People tend to spontaneously name visual stimuli, but they do not spontaneously generate visual images to verbal stimuli, so that visual presentation is more likely to generate two codes than verbal presentation. The existence of two codes could facilitate information processing in addition to augmenting memory. Mental operations like arithmetic are regarded as performed by the articulatory loop. If memory for the stimuli is retained in the visuospatial sketchpad, then the articulatory loop, relieved of memory load, has more capacity for information processing. These findings, if replicated and extended, have broad-ranging implications for education as well as interface design.

Alternatively, it is possible that the greater effort required by aural presentation is due only to differences in the difficulty of perception and not because of any subsequent processing differences, such as visual persistence or differential recruitment of working memory components. Future work could resolve this question by adjusting stimulus discriminability to equalize perception difficulty between the two modes and then check to see whether the effort differences remain.

Further research to determine the true cause of mode-related differences in pupil dilations will help to determine whether such dilations can fulfill Kahneman's second criterion for an effort proxy, intertask comparability, and thus be useful for comparisons of cognitive load between the auditory and visual domains.

#### **REFERENCES**

- Ahern, S., & Beatty, J. (1979). Pupillary responses during information processing vary with scholastic aptitude test scores. *Science*, *205*, 1289–1292.
- Andreassi, J. L. (2006). Pupillary response and behavior. In *Psychophysiology: Human behavior and physiological response* (5th ed, pp. 289–307). Mahwah, NJ: Erlbaum.
- Backs, R. W., & Walrath, L. C. (1992). Eye movement and pupillary response indices of mental workload during visual search of symbolic displays. *Applied Ergonomics*, *23*, 243–254.
- Baddeley, A. D. (2007). *Working memory, thought, and action*. Oxford: Oxford University Press.
- Beaman, C. P. (2002). Inverting the modality effect in serial recall. *Quarterly Journal of Experimental Psychology*, *A55*, 371–389.
- Beatty, J. (1982a). Phasic not tonic pupillary responses vary with auditory vigilance performance. *Psychophysiology*, *19*, 167–172.
- Beatty, J. (1982b). 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.
- Beatty, J., & Lucero-Wagoner, B. (2000). The pupillary system. In J. T. Cacioppo, L. G. Tassinary, & G. Berntson (Eds.), *Handbook of psychophysiology* (2nd ed, pp. 142–162). Cambridge, UK: Cambridge University Press.
- Becker, W., Iwase, K., Jürgens, R., & Kornhuber, H. H. (1976). Brain potentials preceding slow and rapid hand movements. In W. C. McCallum & J. R. Knott (Eds.), *The responsive brain* (pp. 99–102). Bristol, UK: Wright.
- Boersma, F., Wilton, K., Barham, R., & Muir, W. (1970). Effects of arithmetic problem difficulty on pupillary dilation in normals and

- educable retardates. *Journal of Experimental Child Psychology*, 9, 142–155.
- Bradshaw, J. (1967). Pupil size as a measure of arousal during information processing. *Nature*, 216, 515–516.
- Bradshaw, J. L. (1968). Pupil size and problem solving. *Quarterly Journal of Experimental Psychology*, 20, 116–122.
- Bradshaw, J. L. (1969). Background light intensity and the pupillary response in a reaction time task. *Psychonomic Science*, 14, 271–272.
- Bradshaw, J. L. (1970). Pupil size and drug state in a reaction time task. *Psychonomic Science*, 18, 112–113.
- Chen, C. (2004). *Information visualization: Beyond the horizon*. London: Springer Verlag.
- Clark, J., & Paivio, A. (1991). Dual coding theory and education. *Educational Psychology Review*, 3, 149–210.
- Cocker, K. D. (1996). Development of pupillary responses to grating stimuli. *Ophthalmic and Physiological Optics*, 16, 64–67.
- Dabbs, J. M., & Milun, R. (1999). Pupil dilation when viewing strangers: Can testosterone moderate prejudice? *Social Behavior and Personality*, 27, 297–301.
- Fehr, T., Code, C., & Herrmann, M. (2008). Auditory task presentation reveals predominantly right hemispheric fMRI activation patterns during mental calculation. *Neuroscience Letters*, 431, 39–44.
- Goldwater, B. C. (1972). Psychological significance of pupillary movements. *Psychological Bulletin*, 77, 340–355.
- Granholm, E., Asarnow, R. F., Sarkin, A. J., & Dykes, K. L. (1996). Pupillary responses index cognitive resource limitations. *Psychophysiology*, 33, 457–461.
- Greene, R. L. (1992). *Human memory*. Mahwah, NJ: Erlbaum.
- Handy, T. C. (2004). *Event-related potentials: A methods handbook* (1st ed). Cambridge, MA: MIT Press.
- Heitz, R. P., Schrock, J. C., Payne, T. W., & Engle, R. W. (2008). Effects of incentive on working memory capacity: Behavioral and pupillometric data. *Psychophysiology*, 45, 119–129.
- Hess, E. H., & Polt, J. M. (1964). Pupil size in relation to mental activity during simple problem-solving. *Science*, 143, 1190–1192.
- Huynh, H., & Feldt, L. S. (1980). Performance of traditional f tests in repeated measures designs under covariance heterogeneity. *Communications in Statistics: Theory and Methods*, 9, 61–74.
- Jennings, J. R. (1987). Editorial policy on analyses of variance with repeated measures. *Psychophysiology*, 24, 474–475.
- Just, M. A., & Carpenter, P. A. (1993). The intensity dimension of thought: Pupillometric indices of sentence processing. *Canadian Journal of Experimental Psychology*, 47, 310–339.
- Kahneman, D. (1973). *Attention and effort*. Englewood Cliffs, NJ: Prentice Hall.
- Kahneman, D., & Beatty, J. (1966). Pupil diameter and load on memory. *Science*, 154, 1583–1585.
- Kahneman, D., Onuska, L., & Wolman, R. E. (1968). Effects of grouping on the pupillary response in a short-term memory task. *Quarterly Journal of Experimental Psychology*, 20, 309–311.
- Kahneman, D., & Wright, P. (1971). Changes of pupil size and rehearsal strategies in a short-term memory task. *Quarterly Journal of Experimental Psychology*, 23, 187–196.
- Klingner, J., Kumar, R., & Hanrahan, P. (2008). Measuring the task-evoked pupillary response with a remote eye tracker. In *Proceedings of the 2008 Symposium on Eye Tracking Research & Applications* (pp. 69–72). Savannah, GA: ACM.
- Kohn, M., & Clynes, M. (1969). Color dynamics of the pupil. *Annals of the New York Academy of Sciences*, 156, 931–950.
- Libby, W. L., Lacey, B. C., & Lacey, J. I. (1973). Pupillary and cardiac activity during visual attention. *Psychophysiology*, 10, 270–294.
- Loewenfeld, I. (1999). *The pupil: Anatomy, physiology, and clinical applications* (2nd ed., Vol. 1). Oxford, UK: Butterworth-Heinemann.
- Loewy, A. D. (1990). Autonomic control of the eye. In A. D. Loewy & K. M. Spyer (Eds.), *Central regulation of autonomic functions* (pp. 268–285). New York: Oxford University Press.
- Logie, R. H., Gilhooly, K. J., & Wynn, V. (1994). Counting on working memory in arithmetic problem solving. *Memory & Cognition*, 22, 395–410.
- Marshall, S. (2002). The index of cognitive activity: Measuring cognitive workload. In *Proceedings of the 2002 IEEE 7th Conference on Human Factors and Power Plants*, 2002 (pp. 7–5–7–9).
- Miettinen, O., & Nurminen, M. (1985). Comparative analysis of two rates. *Statistics in Medicine*, 4, 213–226.
- Misulis, K. E., & Fakhoury, T. (2001). *Spehlmann's evoked potential primer, 3/e* (3rd ed). Oxford, UK: Butterworth-Heinemann.
- Moresi, S., Adam, J. J., Rijcken, J., Van Gerven, P. W., Kuipers, H., & Jolles, J. (2008). Pupil dilation in response preparation. *International Journal of Psychophysiology*, 67, 124–130.
- Paas, F., & Van Merriënboer, J. (1994). Instructional control of cognitive load in the training of complex cognitive tasks. *Educational Psychology Review*, 6, 351–371.
- Paivio, A. (1990). *Mental representations: A dual coding approach*. Oxford, UK: Oxford University Press.
- Peavler, W. S. (1974). Pupil size, information overload, and performance differences. *Psychophysiology*, 11, 559–566.
- Penney, C. G. (1989). Modality effects and the structure of short-term verbal memory. *Memory & Cognition*, 17, 398–422.
- Pomplun, M., & Sunkara, S. (2003). Pupil dilation as an indicator of cognitive workload in human-computer interaction. In D. Harris, V. Duffy, M. Smith, & C. Stephanidis (Eds.), *Human-Centred Computing: Cognitive, Social, and Ergonomic Aspects*. Vol. 3 of the *Proceedings of the 10th International Conference on Human-Computer Interaction, HCI 2003*, Crete, Greece.
- 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.
- Richer, F., & Beatty, J. (1985). Pupillary dilations in movement preparation and execution. *Psychophysiology*, 22, 204–207.
- Richer, F., Silverman, C., & Beatty, J. (1983). Response selection and initiation in speeded reactions: A pupillometric analysis. *Journal of Experimental Psychology: Human Perception and Performance*, 9, 360–370.
- Sahraie, A., & Barbur, J. L. (1997). Pupil response triggered by the onset of coherent motion. *Graefes' Archive for Clinical and Experimental Ophthalmology*, 235, 494–500.
- Tobii Technologies, Inc. (2007). *Tobii 1750*. Stockholm, Sweden: Author.
- Ukai, K. (1985). Spatial pattern as a stimulus to the pupillary system. *Journal of the Optical Society of America*, A2, 1094–1100.
- Van Orden, K. F., Limbert, W., Makeig, S., & Jung, T. (2001). Eye activity correlates of workload during a visuospatial memory task. *Human Factors*, 43, 111–121.
- Verney, S. P., Granholm, E., & Dionisio, D. P. (2001). Pupillary responses and processing resources on the visual backward masking task. *Psychophysiology*, 38, 76–83.
- Welford, W. T. (1980). *Reaction times*. London: Academic Press.

## SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article:

**Appendix S1.** Code for analyzing data. Source code written in R and Python used to organize, clean, analyze, and graph the data.

**Appendix S2.** Code for collecting data. Source code written in Java used to capture and record the pupil and gaze direction measurements made by the eye tracker and synchronize their timestamps with the timing of stimulus presentation and participants' responses.

**Appendix S3.** Code for presenting stimuli. Source code written in Java used to present the visual and aural stimuli.

**Appendix S4.** Code for running eye tracker. Source code written in C# used to initialize and calibrate the Tobii 1750 eye tracker.

**Appendix S5.** Data. All raw data collected in all experiments, including failed and discarded trials, anonymized.

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.

(RECEIVED August 3, 2009; ACCEPTED March 29, 2010)