Task-Evoked Pupillary Responses, Processing Load, and the Structure of Processing Resources

Jackson Beatty
Department of Psychology and Brain Research Institute
University of California, Los Angeles

A physiological measure of processing load or "mental effort" required to perform a cognitive task should accurately reflect within-task, between-task, and between-individual variations in processing demands. This article reviews all available experimental data and concludes that the task-evoked pupillary response fulfills these criteria. Alternative explanations are considered and rejected. Some implications for neurophysiological and cognitive theories of processing resources are discussed.

That the pupil of the eye dilates during mental activity has long been known in neurophysiology. For example, Bumke, the German neurologist, wrote seven decades ago (as translated in Hess, 1975):

Every active intellectual process, every psychical effort, every exertion of attention, every active mental image, regardless of content, particularly every affect just as truly produces pupil enlargement as does every sensory stimulus. (pp. 23–24)

Only recently has this phenomenon been used as a tool in investigating human cognitive processing. The pupillary dilations that accompany cognitive processes are indeed as pervasive a phenomenon as Bumke had indicated. They occur at short latencies following the onset of processing and subside quickly once processing is terminated. Perhaps most important, the magnitude of pupillary dilation appears to be a function of processing load or "mental effort" required to perform the cognitive task.

These facts led Kahneman (1973) to rely on the task-evoked pupillary response as the primary measure of processing load in his effort theory of attention. He justified the use of this physiological measure in terms of the strong empirical relation between task

demands and pupillary dilation, leading to the conclusion that "the limited capacity and the arousal system must be closely related" (p. 10). Kahneman proposed three criteria for any physiological indicator of 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. The first section of this article reviews the evidence that the task-evoked pupillary response may serve as such an indicator.

Task-Evoked Pupillary Responses as a Measure of Processing Load

Changes in central nervous system activity that are systematically related to cognitive processing may be extracted from the raw pupillary record by performing time-locked averaging with respect to critical events in the information-processing task. A task-evoked pupillary response bears the same relation to the pupillary record from which it is derived as does an event-related brain potential to spontaneous electroencephalographic activity. With averaging, short-latency (onset between 100 and 200 msec), phasic task-evoked dilations appear, which terminate rapidly following the completion of processing.
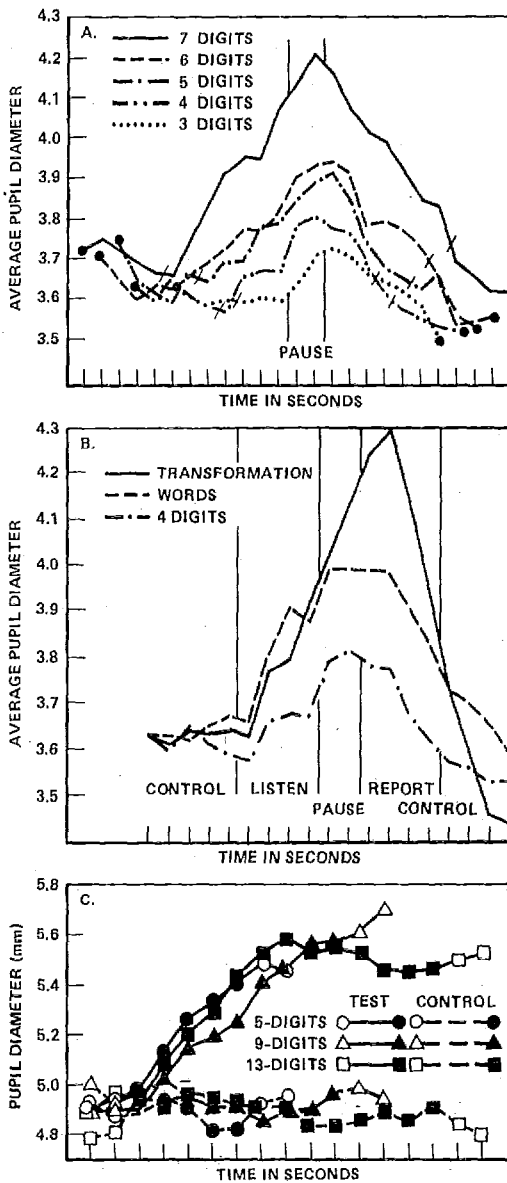
Figure 1. Task-evoked pupillary responses in three short-term memory tasks. [(A) Digit span for string lengths of three through seven digits. Time axes have been adjusted so that the 2-sec pause between presentation and recall is superimposed for all lengths. Slash marks are placed on each curve just before the first digit is presented and after the last digit is repeated. The amplitude of the response grows during item presentation, reaching a maximum during the pause and returns to baseline during report. Peak amplitude is a monotonic function of the number of items held in memory. (Adapted from "Pupil diameter and load on memory" by D. Kahneman & J. Beatty, *Science*, 1966, *154*, 1583–1585. Copyright 1966 by the American Association for

Within-Task Variations in Processing Load

In the last two decades, task-evoked pupillary responses have been obtained for a wide variety of cognitive processes, ranging from sensory detection through memory, language processing, attention, and complex reasoning. The evidence relating to the correspondence between processing demands and pupillary response within individual tasks is as follows.

Short-term memory. The study of short-term memory formed an initial and enduring problem in the pupillometric investigation of information processing. Kahneman and Beatty (1966) presented the first pupillometric analysis of task-evoked pupillary responses in a short-term memory task (see Figure 1A). Strings of three–seven digits were aurally presented at the rate of 1 per sec. After a 2-sec pause, subjects were required to repeat the digit string at the same rate. Under these conditions, pupillary diameter increases with the presentation of each digit, reaching a maximum in the pause preceding report. During report, pupillary diameter decreases with each digit spoken, reaching baseline levels after the final digit. The magnitude of the peak pupillary dilation in this task is an increasing function of string length. Kahneman and Beatty (1966) also observed that if the subject were requested to repeat the string a second time immediately after reporting the final digit, the pupil immediately redilates to the peak diameter for that string and then decreases with each digit spoken until the entire string has been

the Advancement of Science. Reprinted by permission.) (B) Responses for four digits, four words, and a four-digit transformation task. The slope of the pupillary response is a function of item difficulty. (Adapted from "Pupil diameter and load on memory" by D. Kahneman & J. Beatty, *Science*, 1966, *154*, 1583–1585. Copyright 1966 by the American Association for the Advancement of Science. Reprinted by permission.) (C) Pupillary responses during five-, nine-, and 13-digit test and control trials. Pupillary diameter increases until approximately seven items are held in memory, after which the curve of the response becomes asymptotic. (Adapted from "Individual differences in pupil size and performance" by S. Peavler, in *Pupillary dynamics and behavior* by M. Janisse (Ed.), p. 164. Copyright 1974 by Plenum Press. Reprinted by permission.)]

reported for the second time. Beatty and Kahneman (1966) demonstrated that a similar pupillary function is obtained when a string of items is recalled from long-term memory: On request to report, a large pupillary dilation is observed as information is retrieved from long-term memory and the response is organized. As each digit in the string is spoken, pupillary diameter decreases, reaching baseline levels after the last digit is spoken.

The slope of this task-evoked pupillary response is determined by the difficulty of the to-be-remembered information as indexed by memory span for different types of items. Kahneman and Beatty (1966) tested three conditions: recall of four digits, recall of four unrelated nouns, and transformation of a four-digit string by adding one to each item (see Figure 1B). The slope of the pupillary response during input was smallest for the least difficult items, the strings of four digits that were to be simply repeated. A steeper slope was observed for the strings of four words. The greatest slope was obtained for the most difficult task, digit string transformation. Thus, both item difficulty and number of items affect the pupillary response in short-term memory tasks.

The idea that the task-evoked pupillary response provides a physiological measure of processing load received direct support in a subsequent experiment by Kahneman, Beatty, and Pollack (1967), in which both pupillometric and behavioral interference methods were used to assess processing load in the four-digit add-one memory transformation task. In using a secondary task of visual target detection, it was found that the amplitude of the task-evoked pupillary response was a reliable predictor of load-induced performance decrements in the secondary task. A series of controls ruled out any peripheral interference of the pupillary dilations themselves on performance of the secondary task.

Rehearsal strategies that improve performance on a short-term memory task act to reduce the amplitude of the task-evoked pupillary response. Kahneman, Onuska, and Wolman (1968) presented subjects with strings of nine digits, either at a uniform rate of 1 per sec or with a temporally imposed

three-digit grouping (.5 sec separating digits within a group and 2.0 sec separating groups). The grouped mode of presentation had previously been shown to increase digit span materially (Ryan, 1967), presumably by breaking the string into more codable units or chunks (Miller, 1956). The pupillometric data reflected the experimentally induced differences in processing strategy: A steady monotonic increase in pupillary diameter accompanied presentation of the digits at the uniform rate, whereas waves of dilation during presentation and constriction during the intergroup pauses characterized the grouped presentation condition. Thus, the task-evoked pupillary response appears to reflect changes in information-processing demands induced by processing strategies that affect performance.

The idea that the pupillary response measures processing load found further support in Peavler's (1974) study of information overload (see Figure 1C). The capacity of short-term memory for strings of unrelated digits is approximately 7 (Miller, 1956). Peavler measured the task-evoked pupillary response for strings of five, nine, and 13 digits, which were randomly intermixed in presentation. During presentation of the strings, pupillary diameter increased as an increasing function of memory load for digits 1 through 7. At the seventh or eighth digit, the pupillary response reached an asymptote; no further dilation was observed. These data suggest that as long as some information-processing capacity remains, increasing memory load is reflected by increasing pupillary dilation. Once the limits of capacity are exceeded, however, further increases in task demands no longer yield increased pupillary dilation.

Language processing. Several aspects of language processing have been studied pupillometrically. At the most molecular level, Beatty and Wagoner (1978) used an experimental method developed by Posner (Posner & Boies, 1971; Posner & Mitchell, 1967) to study the visual encoding of single letters. In Beatty and Wagoner's first experiment, subjects were required to judge whether or not a pair of visually presented letters had the same name. Individual letters were presented in either upper or lower case type.

Thus, two kinds of letter pairs could be judged to be the same by the name criterion. If both letters are presented in the same case (e.g., AA or aa), only the physical features of the letters need to be analyzed to reach the correct judgment. If they differ in case (e.g., Aa or bB), then, in addition to analyzing the physical features, a second process of name code extraction must be performed. Figure 2A presents these data. Although the task-evoked pupillary responses were small in this simple task (on the order of .1 mm), they did reflect the extra processing required for name code extraction. Significantly larger responses were obtained for letter pairs that differed in case.

In a second similar experiment, Beatty and Wagoner (1978) examined three levels of character encoding by requiring the use of a higher order category classification (vowels and consonants). Thus, letter pairs could be physically identical, identical in name, or identical in category membership (e.g., Ae or BK). Again, the task-evoked pupillary response reflected the processing required to perform the letter-matching task at each level (see Figure 2B).

Ahern (Ahern, 1978; Ahern & Beatty, 1981) undertook two experimental investigations involving language processing as part of a larger research program on individual differences in intelligence. The first of these experiments examined task-evoked pupillary responses in the perception and
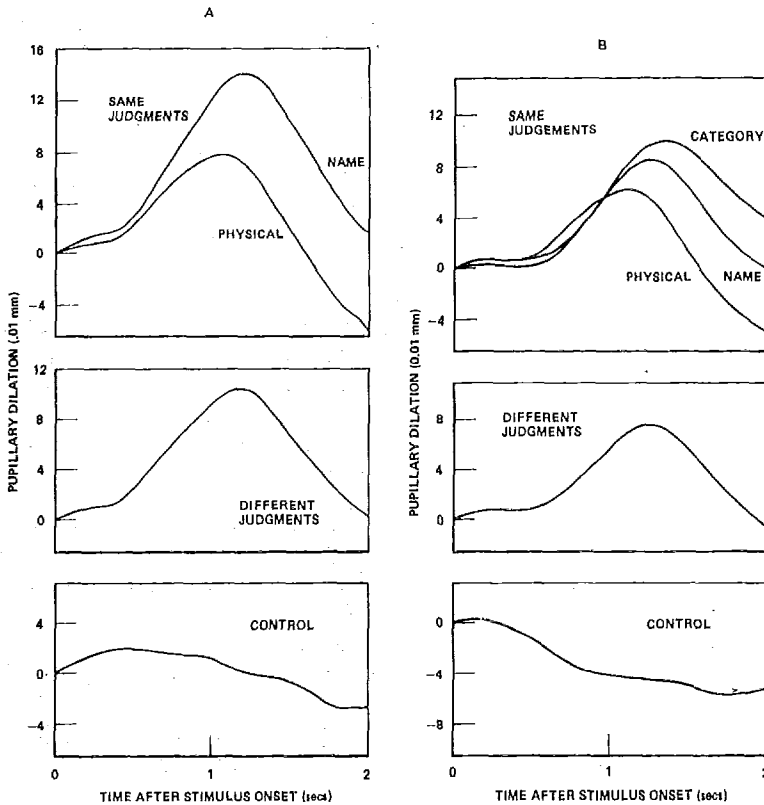

Figure 2. Task-evoked pupillary responses in the Posner letter-matching task. [(A) Responses for correctly identified same and different letter pairs using a name rule. The responses to same judgments are larger when name code extraction is required. In control trials, subjects always saw the letter pair XX and were required to respond "same." (B) Responses in the letter-matching task using a category rule. Again, the amplitude of the responses for same judgments increases with complexity of processing required to reach that judgment. (Adapted from "Pupillometric signs of brain activation vary with level of cognitive processing" by J. Beatty & B. L. Wagoner, Science, 1978, 199, 1216–1218. Copyright 1978 by the American Association for the Advancement of Science. Reprinted by permission.)]
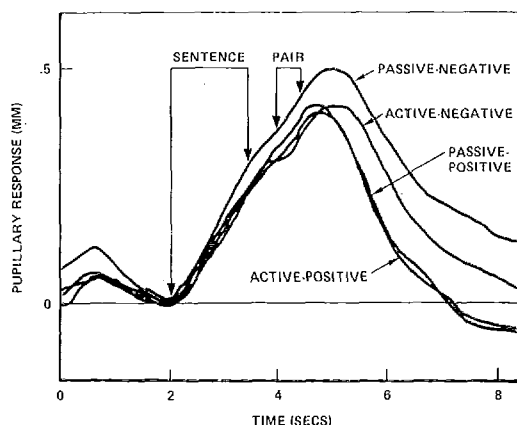
Figure 3. Task-evoked pupillary responses for four levels of sentence complexity in Baddeley's Grammatical Reasoning Task. [The amplitude of the response is significantly greater for the longer, syntactically more complex sentences. (Adapted from Ahern, 1978).]

comprehension of words. Subjects were required to judge pairs of words as similar or different in meaning. The first word of each pair was drawn from either the easiest or the most difficult items of one of three psychometric vocabulary tests. The second word, presented 2 sec later, was either a synonym of the first or quite different in meaning. In this experiment, the ease of retrieving lexical information was reflected in the pupillary response. A dilation of approximately .1 mm followed the presentation of the easy target words, whereas the dilation for the difficult target words was twice as large. A second dilation followed the presentation of the comparison word, yielding pupillary dilations of .30 and .35 mm, respectively, during the judgment period. Thus, larger pupillary dilations accompany the semantic processing of psychometrically more difficult vocabulary items.

At the most molar level, task-evoked pupillary responses have been studied as subjects processed meaningful sentences of different complexity. Ahern (Ahern, 1978; Ahern & Beatty, 1981), using Baddeley's Grammatical Reasoning Task (Baddeley, 1968), presented sentences of the form "A follows B" or "B precedes A," after which an exemplar "AB" or "BA" was given. The task was to determine whether the sentence correctly described the exemplar. Sentences

differed in grammatical complexity, being active–positive, active–negative, passive-positive, or passive–negative. Although these sentences differed in length, sentence duration was held constant by using computer presentation of digitized natural speech. In this task, increasing dilation was observed during the presentation of the sentence and the exemplar, which peaked during the decision interval (see Figure 3). The amplitude of these responses averaged approximately .40 mm and differed significantly as a function of grammatical complexity, with the longer, more complex sentences eliciting larger pupillary responses.

Wright and Kahneman (1971) also applied pupillometric measurements in a sentence-processing task. Subjects were presented with complex sentences of the form "The qualified managing director was recently sensibly appointed by the expanding successful company." Subjects were required either to repeat the sentence or to answer a question of the form "Who appointed the director in this sentence?" The query was posed either before or after the sentence was presented. When the task was to repeat the sentence, the task-evoked pupillary response increased as the sentence was presented and peaked during the retention interval (3 or 7 sec), reaching a maximum dilation of approximately .30 mm. When the question was posed after the pause, peak dilation during the pause was approximately .20 mm and was followed by another dilation as the answer to the question was formed. The peak of this dilation was approximately .40 mm with respect to pre-sentence baseline. When the question was posed before sentence presentation, the task-evoked pupillary response rose more gradually but increased rapidly when the relevant portion of the sentence was presented, indicating organization and processing of the answer to the query. No evidence of processing of phrase boundaries was observed, but, as Wright and Kahneman commented, their sentences were not representative of those naturally occurring in spoken English.

Beatty and Schluroff (Note 1) studied the effects of both syntactic and semantic organization on the task-evoked pupillary re-

sponse and performance in the encoding and reproduction of six-word sentences. A base set of 12 standard sentences using one of six syntactic constructions was used to test performance with both syntactic and semantic organization. An example of such a sentence is "Should blind people lead quiet lives?" In a second condition, the semantic organization of these sentences was reduced by exchanging words between sentences while maintaining the syntactic frame (e.g., "Many blind roses play heavy trouble"). In the third condition syntactic organization was also eliminated by selecting random strings of items (e.g., "Rains children milk golden usually medals"). Figure 4 presents the task-evoked pupillary responses obtained under these conditions. First, on each of the records, small dilations may be observed during the reception and production of each of the individual words in the string. Second, both syntactic and semantic organization acted to reduce the amplitude of the task-evoked pupillary response. Third, these effects of linguistic organization were present at input

and during the pause preceding report: No additional effect of linguistic organization was observed during production. Finally, Beatty and Schluroff conducted a secondary analysis for the two conditions with syntactic organization in which sentence frames of lower and higher complexity were compared. The more complex syntactic frames yielded significantly larger task-evoked pupillary responses in both normal and semantically anomalous sentences.

Reasoning. Mental arithmetic has been used as an example of a complex reasoning problem by several investigators. Hess and Polt (1964), in their initial and influential article on pupillary signs of mental activity, measured pupillary diameter as five subjects solved four multiplication problems, ranging in difficulty from 7×8 to 16×23. For each of the subjects and each of the problems, pupillary diameter increased from the moment of problem presentation until the point of solution. Hess and Polt reported these data as percentage dilation, not as absolute values. Across subjects, the percentage di-
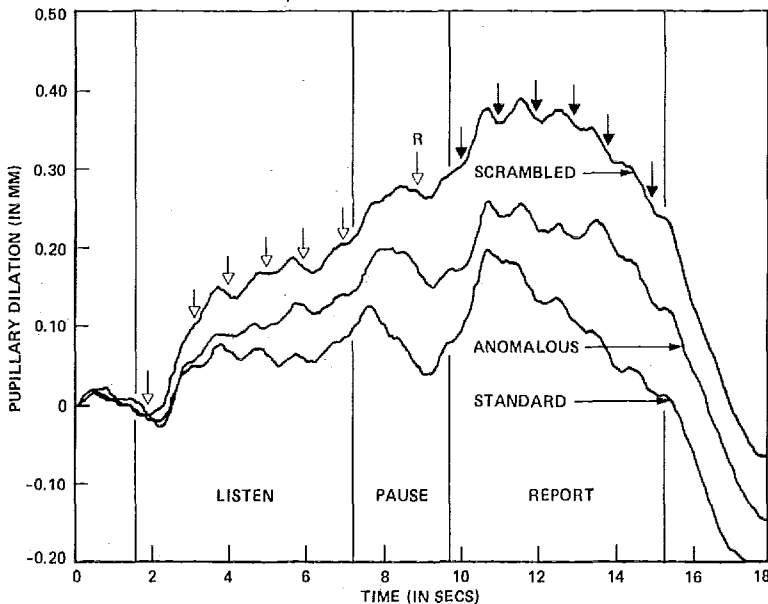
Figure 4. Task-evoked pupillary responses for six-word sentences differing in linguistic organization. [Standard sentences were meaningful English sentences. Anomalous sentences used the same syntactic frames but with words interchanged between sentences to render the strings nearly meaningless. Scrambled sentences had neither syntactic nor semantic organization. Both syntactic and semantic organization independently reduced the processing load imposed by the sentence repetition task. Open arrows indicate presentation of words and the response cue; filled arrows indicate timing clicks. (From Beatty & Schluroff, Note 1).]

lation was perfectly ordered by difficulty of the problem.

Bradshaw (1968) reported similar results for six subjects performing mental division problems at two levels of difficulty. Pupillary diameter increased during problem solving until the point of solution, peak dilation being larger for the more difficult problems. Similarly, Payne, Parry, and Harasymiw (1968) described a monotonic relation between mean pupillary diameter and problem difficulty but noted that this relationship is markedly nonlinear with respect to difficulty scales based on percent correct solution, time to solution, or subjective rating of difficulty. Pupillary diameter in mental multiplication appears to peak rapidly as a function of difficulty, with more difficult problems requiring more time until solution is reached.

These results were subsequently replicated by Ahern and Beatty (1979, 1981). Three levels of problem difficulty were used, ranging from multiplying pairs of one-digit numbers to multiplying pairs of two-digit numbers ranging between 11 and 20. Figure 5 presents these data. In this task an initial dilation of approximately .15 mm accompanies the encoding and storage of the multiplicand. The second and major dilation follows presentation of the multiplier and continues through problem solution. Both the amplitude and latency of this latter dilation increase as a function of problem difficulty. In the most difficult condition, the response reached an asymptote at approximately .50 mm.

Perception. Small but reliable pupillary dilations accompany the detection of both visual and acoustic signals at near-threshold intensities. Hakerem and Sutton (1966) provided the first pupillometric analysis of processing load in perceptual detection. Subjects viewed a uniform visual field on which brief increments in luminance could be imposed as pupillary diameter was measured. When the magnitude of the intensity increment was adjusted to yield 50% correct detection, all vestiges of the flash-induced light reflex disappeared. Under these conditions, a clear pupillary dilation of approximately .10 mm was observed if, and only if, a presented target was detected. Figure 6A presents these data.

Beatty and Wagoner (Note 2) extended Hakerem and Sutton's (1966) finding to audition, using weak 100 msec 1 kHz sinusoidal acoustic signals presented against a background of white noise. Signals were presented on each trial with a probability of .50. After each trial, the subjects rated their certainty that a target had or had not been presented (Green & Swets, 1966). For signal-present trials, the magnitude of the task-evoked pupillary response was largest for signals judged with high certainty to be present and smallest for signals judged with high certainty to be absent. Amplitudes for uncertain judgments assumed intermediate values (see Figure 6B). These results fully confirm those reported by Hakerem and Sutton for visual detection.

It is of interest that the signal detection task provides one instance in which increasing task difficulty does not increase the amplitude of the pupillary response. Beatty and Parasuraman (Note 3) reported that manipulation of the signal-to-noise ratio of the stimulus affects the performance but not the amplitude of the task-evoked pupillary response. They interpret this finding as further evidence that acoustic signal detection is a data-limited, not a resource-limited, process (Norman & Bobrow, 1975). For a data-limited process, performance quality is determined by input data quality; there are no additional processing steps or procedures that the listener may use to increase further his or her ability to detect the signal. It therefore follows that decreasing the detectability of a weak sensory signal aversely affects performance without increasing the processing load associated with the task.

Task-evoked pupillary responses are also observed in perceptual discrimination tasks, in which a presented stimulus must be compared against memory and a judgment rendered. Kahneman and Beatty (1967) reported the first study of the pupillary response in perceptual discrimination. On each trial the subject heard a standard tone of 850 Hz, which was followed 4 sec later by a comparison tone. The comparison was one of 11 frequencies, ranging between 820 and 880 Hz in 6-Hz steps. The subject's task was to judge whether the comparison tone was higher or lower in pitch than the standard
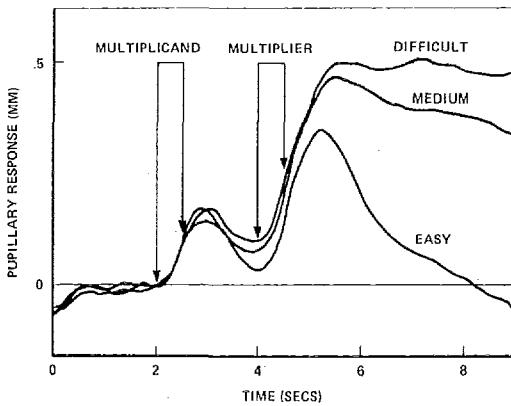
Figure 5. Task-evoked pupillary responses in mental multiplication. [There is an initial dilation during the encoding of the multiplicand followed by a substantial dilation on presentation of the multiplier and the beginning of problem solving. Both the amplitude and the peak latency of this major dilation increase as a function of problem difficulty. (From Ahern, 1978).]

dropped as a function of time over the 48 min of the task (from 84% to 67%). The amplitude of the task-evoked pupillary response showed a similar reduction, from ap-

tone. The difficulty of the discrimination is inversely related to the difference in pitch between the comparison and the standard. The amplitude of the response to the comparison tone varied as a function of discrimination difficulty, from approximately .10 mm for easy to .20 mm for difficult discriminations.

Sustained attention. Processing efficiency in memory-dependent perceptual discrimination tasks is known to deteriorate if the task is prolonged and the number of discriminations required per minute is relatively high (Parasuraman, 1979; Parasuraman & Davies, 1977). One theory to explain this vigilance decrement is that central nervous system activation deteriorates over time under such conditions, and as a result the adequacy of information processing is increasingly compromised. Such changes might appear in either tonic or phasic pupillometric measures. Thus, Beatty (in press) measured task-evoked pupillary responses to nontarget stimuli in an auditory vigilance task. Nontarget stimuli were 50-msec 1-kHz tone bursts, presented at intervals of 3.2 sec. Randomly intermixed were target stimuli, which were attenuated by 3.5 db. Subjects reported the detection of targets by depressing a microswitch. Under these conditions, the efficiency of target/nontarget discrimination

Figure 6. Task-evoked pupillary responses in signal detection tasks. [(A) Visual signal detection. A clear .1-mm dilation is observed following detection of a signal that is not present for undetected signals or blank trials. (From "Pupillary response at visual threshold" by G. Hakerem & S. Sutton, *Nature,* 1966, *212*(5061), 485–486. Copyright 1966 by Macmillan Journals Limited. Reprinted by permission.) (B) Auditory signal detection. When a rating response (YC = yes, certain; YU = yes, uncertain; NU = no, uncertain; NC = no, certain) is used, the pupillary response increases with the judged likelihood that a signal was presented on that trial. (From Beatty & Wagoner, Note 2).]

CHANNEL ATTENDED

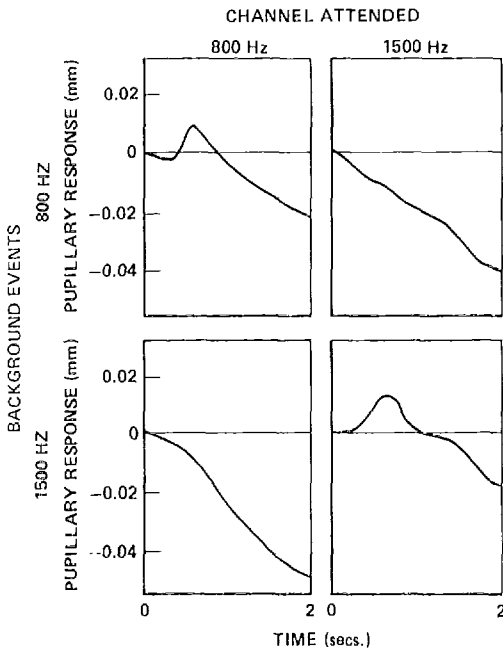*Figure 7.* Task-evoked pupillary responses in auditory selective attention. [Small (.01 mm) dilations may be observed following presentation of background stimuli on the attended but not on the unattended auditory channel. (From Beatty, Note 4).]

proximately .07 mm in the first third of the task to .04 mm in the last. Tonic or baseline pupillary diameter exhibited no such relation with performance. Thus, the physiological mechanisms operating during information processing under alerted conditions appear to be altered under conditions eliciting a vigilance decrement.

Selective attention. Selective attentional processing of sensory information occurs under conditions of high information load when it is not possible to process adequately all incoming information. A commonly cited example of selection in linguistic processing is the cocktail party phenomenon, in which the listener selects one voice among many to be attended to and processed. Electrophysiological evidence of selective attentional processes has been obtained by Hillyard and his co-workers (Hillyard, Hink, Schwent, & Picton, 1973) using a multiple channel tone discrimination task. Beatty (Note 4) used this procedure to test for the effects of selective attention in the task-

evoked pupillary response. Subjects were required to listen to a series of randomly presented high- and low-frequency tones presented at an average rate of 3 per sec. The subject's task was to attend to one of the two types of tones and to press a switch whenever a target tone (marked by a slight frequency increment) occurred. Under these conditions a small (.015 mm) pupillary dilation at a latency of 600 msec followed presentation of nontarget tones on the attended channel, which was completely absent following stimuli on the nonattended channel. Figure 7 presents these data. Although the amplitude of these responses was extremely small, the differences between attended and nonattended tones was highly significant.

Between-Task Variations in Processing Load

Kahneman's (1973) second criterion for a physiological measure of mental effort is that the measure should order variations in processing demands across qualitatively different mental tasks. In each of the experiments described in the preceding section, there appears to be an orderly relationship between the processing demands imposed by a cognitive task and the amplitude of the task-evoked pupillary response. Moreover, tasks that place large demands on the information-processing system—judged behaviorally, subjectively, or by an analysis of task requirements—elicit large task-evoked pupillary responses. Less demanding tasks elicit smaller responses. It is possible, therefore, that task-evoked pupillary responses associated with cognitive function might provide a common metric for the assessment and comparison of information-processing load in tasks that differ substantially in their functional characteristics. Underlying this proposal is the idea that the dynamic changes indexed by the task-evoked pupillary response reflect a basic physiological aspect of processing load that is independent of qualitative differences between tasks.

The usefulness of such intertask comparisons is strengthened by the finding that the magnitude of the task-evoked pupillary responses during cognitive processing is independent of baseline pupillary diameter over

a physiologically reasonable but not extreme range of values (Bradshaw, 1969, 1970; Kahneman & Beatty, 1967; Kahneman et al., 1967). It is therefore possible to compare the absolute values of the task-evoked dilations reported from different laboratories for qualitatively different tasks. Figure 8 presents such a quantitative comparison, giving the approximate peak amplitude of the task-evoked pupillary response measured from published figures for each of the tasks detailed above, subject only to the constraint that the data are not confounded by the effects of overt response.

The leftmost panel of Figure 8 presents peak dilations for short-term memory tasks. The data for short-term retention of digits are the average of the values obtained by Ahern (1978), Kahneman and Beatty (1966), Kahneman, Onuska, and Wolman (1968), and Peavler (1974). The value for retention of four words is from Kahneman and Beatty. The next panel summarizes the literature on language processing. The peak value for the letter-matching task (Posner & Mitchell, 1967) is the average of both experiments published by Beatty and Wagoner (1978). Sentence encode-1 is from Wright and

Kahneman (1971). Sentence encode-2 is from Beatty and Schluroff (Note 1). All other values for language-processing tasks are taken from Ahern. Word encoding is the response to the presentation of the first word in the synonyms judgment task. The values for easy and difficult word matching are the peak response during the judgment period following presentation of the second word in that task. The value for grammatical reasoning is the average of the four types of sentences in Baddeley's Grammatical Reasoning Task (Baddeley, 1968).

The third panel presents data from the mental multiplication task used as an example of complex reasoning. Only Ahern (1978) presented task-evoked pupillary responses for this task that are necessary for comparative peak measurement. Multiplicand storage is the amplitude of the peak response to the first item in the multiplication task. The other three values are the peak amplitudes attained during problem solution.

The rightmost panel presents data for perceptual tasks. The visual detection data are from Hakerem and Sutton (1966), and the auditory detection data are from Beatty and
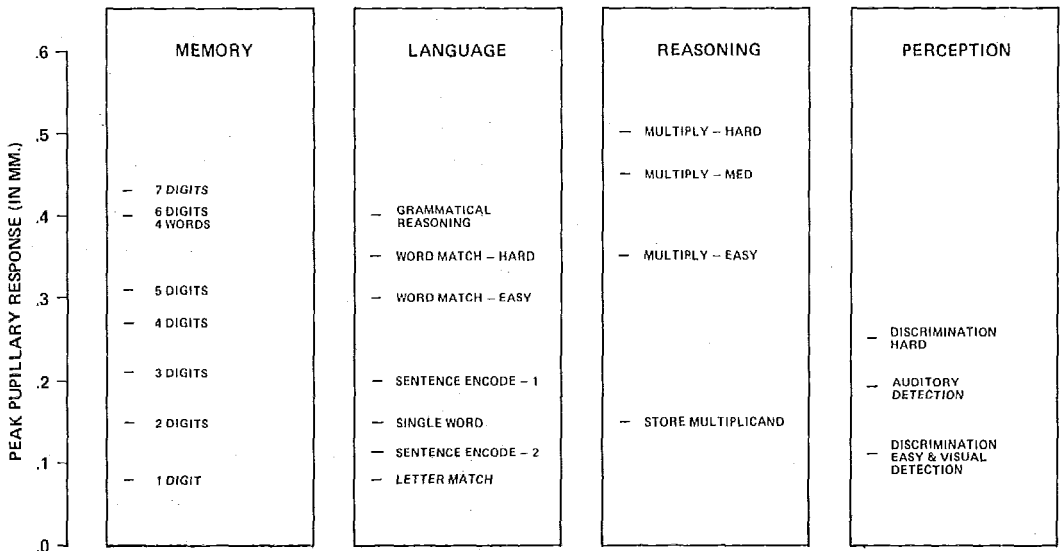

Figure 8. Peak amplitudes of the task-evoked pupillary responses obtained in a range of qualitatively different cognitive tasks, arranged by type of task. (The pupillary response provides a reasonable ordering of tasks on the basis of presumed processing load. See text for further details.)

Wagoner (Note 2). The discrimination data are taken from Kahneman and Beatty (1967).

Several points concerning these data deserve mention. First, the data are tolerant of the stringent demands placed on them in comparing absolute dilation values across experiments. Rescaling of some sort is often. required for physiological data to remove individual differences in responsivity (Johnson & Lubin, 1972). No such rescaling was undertaken here. The data plotted are absolute peak dilations obtained from different groups of subjects performing a range of cognitive tasks under varying experimental conditions in different laboratories. Second, the data plotted in Figure 8 are internally consistent. No abnormally large or small values are present. Third, the ordering of these values corresponds quite closely to an ordering of these tasks using other criteria of information-processing load. The short-term memory tasks cover a large range of values, depending on the number of items held for recall. Similarly in language processing, the sentence comprehension tasks yield large pupillary dilations, whereas the simpler word- and letter-matching tasks elicit much smaller values. The mental multiplication tasks again span a wide range of values, each appropriate to the difficulty of the particular problem. Finally, the perceptual tasks, which behavioral techniques indicate impose minimal processing load, are associated with small task-evoked pupillary responses (Wickens, Note 5).

Thus, Figure 8 provides evidence that the task-evoked pupillary responses faithfully reflect variations in processing load between qualitatively different cognitive tasks. In fact, this physiological phenomenon provides a primary reason for retaining some form of a general metric of processing load, an idea that has recently come under attack because of a failure of simple general capacity models to predict adequately two-task interactions when time sharing (Navon & Gopher, 1979). Thus, the task-evoked pupillary response seems to fulfill Kahneman's second criterion of measuring variations of processing load between qualitatively different mental tasks.

Between-Individual Variations in Processing Load

The only published test of task-evoked pupillary responses as an index of between-subject variations of processing load imposed by a cognitive task was provided by Ahern and Beatty (1979, 1981). They measured task-evoked pupillary responses in two groups of university undergraduates who differed in psychometrically measured intelligence. Subjects were selected on the basis of combined verbal and quantitative Scholastic Aptitude Test scores, being either under 950 or above 1350 for the low- and high-intelligence groups, respectively. Four cognitive tasks were used: mental multiplication, digit span, vocabulary information, and sentence comprehension. In each of these tasks, at least two levels of task difficulty were used, and in all cases the more difficult task parameters elicited larger task-evoked pupillary responses. Further, all tasks were sensitive to the between-group differences in putative intelligence; in each task, subjects in the high-intelligence group made fewer errors.

For three of the four tasks, significant between-group differences in the amplitude of the task-evoked pupillary response were observed. Figure 9 presents the task-evoked pupillary responses for the mental arithmetic task. With the exception of the vocabulary task, in which the pupillary responses were essentially identical in both groups, the task-evoked pupillary response amplitudes were consistently smaller for the more intelligent subjects than for their less intelligent counterparts. These between-group differences were interpreted as indicating that performance of the same objective cognitive tasks is less demanding for more intelligent individuals. In addition, the amplitudes of the autonomically mediated light and dark reflexes were measured to test for possible confounding with group differences in autonomic excitability. The reflex responses, however, were identical in the two groups, suggesting that the observed between-individual differences in the amplitude of the task-evoked pupillary response reflect central rather than peripheral aspects of neural

function. These data suggest that the task-evoked pupillary response might fulfill Kahneman's third requirement for a measure of processing load, that the index would reflect between-individual differences as well as task differences.

Alternative Interpretations of the Task-Evoked Pupillary Response

The definition of the task-evoked pupillary response as a measure of processing load depends not only on the clear demonstration that the response varies with relevant task parameters but also on evidence excluding the involvement of other potentially confounding variables. In this context, it is important to distinguish between factors affecting the task-evoked pupillary response and those that affect tonic or baseline pupillary diameter. Basal diameter is strongly influenced by a wide variety of systemic and environmental factors (Lowenstein & Lowenfeld, 1962). The procedure of averaging phasic changes in pupillary diameter with respect to a significant event in the experimental trial, however, ensures that general factors cannot affect the task-evoked pupillary response, except as they may systematically vary during the course of an experimental trial. Nonetheless, there has been some concern that a portion of the variance of the task-evoked pupillary response may be attributed to noncognitive variables, particularly to the light reflex and to emotional processes (Goldwater, 1972).

The possibility that the light reflex might affect the task-evoked pupillary response must be considered whenever subjects are permitted to shift their gaze across a non-uniform visual field, such as a photograph (Hess, 1975). The controlled use of a fixation point and the use of constant illumination visual displays, however, render interpretations of phasic changes as reflections of the light reflex highly unlikely. Visual and oculomotor factors may be reasonably discounted when subjects are required to maintain fixation and all experimental stimuli are aurally presented.

The question of emotional factors exerting systematic influence on the phasic task-evoked pupillary responses is more complex, particularly as such factors are often only poorly and intuitively defined. In the context of the cognitive experiments described above, emotional involvement might be expected to
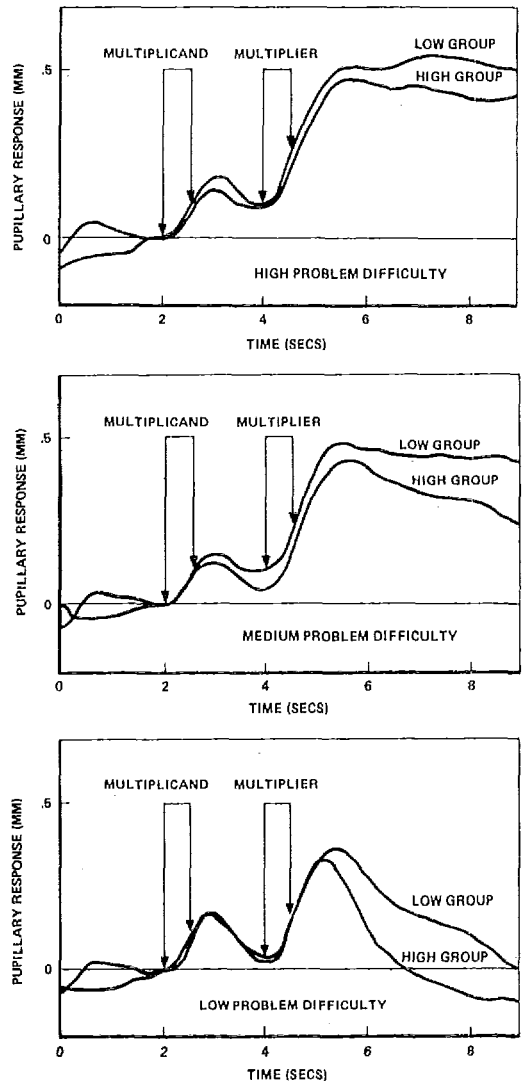
Figure 9. Task-evoked pupillary responses for correctly solved problems at three levels of difficulty for subjects of high and low psychometrically defined intelligence. [At all difficulty levels, larger pupillary responses were observed for subjects in the lower intelligence group. (From "Pupillary responses during information processing vary with Scholastic Aptitude Test scores" by S. Ahern & J. Beatty, *Science,* 1979, *205,* 1289–1292. Copyright 1979 by the American Association for the Advancement of Science. Reprinted by permission.)]

manifest itself in terms of task-induced anxiety reactions. Several lines of reasoning tend to discount such arguments.

First, task-evoked pupillary responses have been reliably observed in tasks in which it is difficult to hypothesize emotional involvement. For example, in an auditory selective attention task (Beatty, Note 4), small but consistent task-evoked pupillary responses of approximately .015 mm were observed following presentation of nonsignal tone on the attended channel, whereas no responses were present to tones on the unattended channel. Because the event rate in that experiment was 3 per sec, an average of 1.5 dilations was obtained each second for the duration of the testing procedure (approximately 15 min). Considering these dilations to be a long string of stimulus selective, high-speed emotional reactions strains the concept of emotional response to a point of absurdity. Similar arguments might be made for a variety of other simple cognitive tasks that would not appear to arouse emotion or to induce anxiety for any subject.

A second reason to reject an emotion hypothesis as an explanation of task-evoked pupillary responses is based on an investigation of individual differences in pupillary response amplitude. Ahern (Ahern, 1978; Ahern & Beatty, 1979, 1981) obtained pupillary responses in 39 university undergraduates tested in four cognitive tasks. There was a significant correlation between a psychometric measure of subject intelligence and the amplitude of the task-evoked pupillary responses in the cognitive tasks. There was no significant correlation, however, between the amplitude of the pupillary response and either state or trait anxiety (Spielberger, 1968). Differences in amplitude of the task-evoked pupillary response between individuals appears to be a function of differences in cognitive ability rather than emotionality.

Third, in his study of information overload in the digit span task, Peavler (1974) also addressed the question of interpreting pupillary dilation as an indication of emotional factors. Peavler reasoned that his data are incompatible with any interpretation of the task-evoked pupillary response as a reflec-

tion of task anxiety or other emotional responses to the testing situation. If the task-evoked pupillary response reflected emotional factors due to fear of performance failure, then a large dilation should accompany the presentation of the later digits in the 13 digit strings because only at this time could the subjects know that the limits of capacity would be exceeded and that their performance could not be perfect. No such dilations to information overload were observed.

These lines of argument suggest that emotional factors are relatively unimportant as determinants of the pupillary responses observed in carefully controlled information-processing tasks. Although emotional factors are well known for their expression in the autonomic nervous system, the effects of emotional arousal are generally longer lasting than the brief phasic responses evoked by cognitive activity (Lang, Rice, & Sternbach, 1972). Thus, changes in emotionality are more likely to affect the tonic or basal pupillary diameter and not the phasic responses under discussion here.

Several other types of potentially confounding variables have also been tested. The effects of motivation on the pupillary response were tested by Kahneman, Peavler, and Onuska (1968, Experiment 2) in a short-term memory task by varying monetary incentives associated with correct performance on different trials. Increasing the incentives had no effect on performance, nor did it affect the task-evoked pupillary response during the performance of either of the short-term memory tasks.

Clark and Johnson (1970) tested the possibility that the task-evoked pupillary response in short-term memory experiments might come from the subject's knowledge about the results of previous pupillometric studies and the demand characteristics of the experiment. Varying these expectations had no effect on the pupillary response, which conformed to the pattern previously reported by Kahneman and Beatty (1966).

Taken together, these lines of evidence lend support to the original hypothesis of Kahneman and Beatty (1966) that the task-evoked pupillary response reflects the mo-

mentary level of processing load and is not an artifact of noncognitive confounding factors.

Processing Load and Processing Resources

Kahneman, in his 1973 theory, identified processing load with amount of attention or "mental effort" allocated to the performance of a mental operation or task. This type of capacity theory, like that previously suggested by Moray (1967), proposed that more than one mental operation might be simultaneously executed providing that the joint demand for attention does not exceed the available supply of attention or processing capacity. The only restriction placed on this prediction was that the two operations do not require the simultaneous use of fixed processing structures, such as sensory or motor channels.

Performance data obtained with time-shared mental tasks, however, provide little support for any general capacity model. Apparently trivial changes in the structure of a task may produce large differences in its interaction with other tasks, even though the information-processing characteristics of the task remain unaltered. Further, when task difficulty levels are varied, some pairs of tasks show performance interactions, whereas others do not (Wickens, 1979). These findings have led most theorists toward multiple capacity models (Navon & Gopher, 1979, 1980; Sanders, 1979; Wickens, 1979, 1980). These models postulate several types of processing capacity that may be allocated among mental operations. In such formulations, mental operations may be performed simultaneously without interference if the demand for capacity from any of the multiple pools of capacity does not exceed the capacity available in each pool. These separate, qualitatively distinct types of information-processing capacity are commonly called processing resources (Norman & Bobrow, 1975). In a multiple resource model, processing load reflects the aggregate-processing resources consumed in the performance of a mental operation.

The major issue confronting multiple resource theory is the identification and specification of these specialized processing resources. This had proved exceedingly difficult. Although it has been proposed that resources may be identified from task interactions alone (Navon & Gopher, 1979), most theorists have sought other sources of converging information in attempting to identify these specific processing capacities.

Kinsbourne (Kinsbourne & Hicks, 1978), for example, suggested that spatially restricted, functionally specialized regions of the cerebral cortex constitute the processing resources of multiple resource theory. Kinsbourne and his collaborators presented a substantial amount of evidence suggesting that the interference between two mental operations increases as a function of the spatial proximity of the primary cortical regions involved in performing those operations. A related view has been proposed by Wickens (1980), who suggested that processing resources may be categorized by input modality (visual or auditory), hemispheric processing specialization (spatial or verbal), and type of responding (vocal or manual). Restated in terms of cortical regions, the auditory and visual sensory cortices, the association cortices of the right and left hemisphere, and the highly differentiated hand and mouth-respiratory regions of the motor and premotor cortex constitute putative processing resources in Wickens's model. The view that the multiple, functionally specialized processing resources of the human information-processing system may be identified with functionally specialized regions of the human cerebral cortex is a sensible suggestion.

Nevertheless, if restricted regions of the cerebral cortex in fact form the resources of the human information-processing system, why should the task-evoked pupillary response reflect their utilization? The answer seems to lie in the dynamic interaction between cerebral cortex and the reticular activating system of the brainstem, coupled with the fact that pupillary movements provide a sensitive indicator of reticular function. Luria provided a very clear overview of the reciprocal interactions of cerebral cortex and the reticular core in his popular 1973 monograph:

[The] maintenance of the optimal level of cortical tone is essential for the organized course of mental activity. . . . [However], the structures maintaining and regulating cortical tone do not lie in the cortex itself, but below it [in the reticular formation of the brainstem]. . . . Some of the fibres of [the] reticular formation run upwards to terminate in higher nervous structures such as the thalamus, caudate body, archicortex and, finally, the structures of the neocortex. . . . [They play] a decisive role in activating the cortex and regulating the state of its activity. (pp. 45–46)

The higher levels of the cortex, participating directly in the formation of intentions and plans, recruit the lower systems of the reticular formation of the thalamus and brain stem, thereby modulating their work and making possible the most complex forms of conscious activity. (p. 60)

Task-evoked pupillary dilations very likely reflect the cortical modulation of the reticular core during cognitive processing. The pupillary system, like other peripheral systems, receives input from most structures in the reticular formation. It must be remembered that the efferent fibers leaving reticular structures typically bifurcate, sending one branch upward to the forebrain and another downward, synapsing on a wide variety of motor nuclei (Brodal, 1981). Thus, any response to forebrain commands modulating activity in the cortico-reticular reticulo-cortical loop will also make its effects felt in the autonomic periphery. For this reason, pupillary movements have served neurophysiology well as a sensitive indicator of reticular system discharge (Moruzzi, 1972).

The evidence reviewed above argues that the amplitude of task-evoked pupillary responses provides a reliable index of task-induced processing load. The use of time-locked averaging methods ensures that only changes related in time to the mental operations under study are measured. These responses grow larger as task parameters are varied to increase task demands for processing resources (Navon & Gopher, 1979). Moreover, the measure reflects variations in processing load between qualitatively different mental operations in a reasonable and consistent manner. For this reason, task-evoked pupillary responses may provide a global indication of task-induced processing load even when the composition of processing resources differs between tasks. There is nothing incompatible in viewing the pupillary response as a measure of the aggre-

gate task-induced utilization of multiple processing resource. This idea is in some ways analogous to the use of a general physiological measure such as oxygen uptake as an indicator of the aggregate metabolic demands of a set of functionally distinct organs.

The task-evoked pupillary response may also provide an indication of the joint demand for resources in pairs of time-shared tasks. No experiments explicitly testing this two-task prediction, however, have yet been published. Nevertheless, some data obtained by Kahneman, Peavler, and Onuska (1968, Experiment 1) on the effect of motor responding and cognitive processing are relevant here. Kahneman et al. examined the effects of verbalization on the task-evoked pupillary response in a short-term memory task at two levels of difficulty. Subjects listened to a string of four digits that they were to repeat or transform by adding one (Kahneman & Beatty, 1966). After presentation, they either repeated the appropriate response twice at the rate of 1 digit per sec or mentally produced the response in the first interval and verbally produced it in the second. The more difficult digit transformation task yielded systematically larger pupillary dilations regardless of verbalization condition. The form of the response was unaltered in the absence of verbalization. The effect of verbalization was to increase the amplitude of the task-evoked pupillary response. These data are consistent with the hypothesis that the organization of an overt motor act places additional demands on information-processing resources that are reflected in the task-evoked pupillary response. This finding indicates that a physiological approach to the study of dual task interactions is possible.

Finally, it is interesting to note that, although pupillometric data formed the basis for Kahneman's general capacity theory, in their original research reports Kahneman and Beatty (1967) proposed a concept of processing load suggestive of modern multiple resource theories:

The frequent use of the concept of processing load in the present paper has been guided by a simple analogy: consider a houseful of electrical devices that are variously put in operation by manual switches or by their

own internal governor systems. The total amperage demanded by the entire system at any one time may easily be read on an appropriate electrical instrument outside the house. Processing load is here construed as analogous to an aggregate demand for power, and there is ground for the hope that the pupil may function as a useful approximation to the relevant measuring device. (p. 104)

The task-evoked pupillary response does serve as such a measure. It provides a reliable and sensitive indication of within-task variations in processing load. It generates a reasonable and orderly index of between-task variations in processing load. It reflects differences in processing load between individuals who differ in psychometric ability when performing the same objective task. For these reasons, the task-evoked pupillary response provides a powerful analytic tool for the experimental study of processing load and the structure of processing resources.

Reference Notes

1. Beatty, J., & Schluroff, M. *Pupillometric signs of brain activation reflect both syntactic and semantic factors in language processing.* Paper presented at the meeting of the Society for Psychophysiological Research, Vancouver, October 1980.
2. Beatty, J., & Wagoner, B. L. *Activation and signal detection: A pupillometric analysis* (Tech. Rep. 10). Los Angeles: University of California, Los Angeles, Human Neurophysiology Laboratory, August 1977.
3. Beatty, J., & Parasuraman, R. *Task-evoked pupillary responses in auditory signal detection.* Unpublished manuscript, 1981. (Available from J. Beatty, Department of Psychology, University of California, Los Angeles, California 90024).
4. Beatty, J. *Pupillometric signs of selective attention in man* (Tech. Rep. No. 18). Los Angeles: University of California, Los Angeles, Human Neurophysiology Laboratory, February 1980.
5. Wickens, C. D. *The effects of time sharing on the performance of information processing tasks: A feedback control analysis* (Tech. Rep. No. 51). Ann Arbor: University of Michigan, Human Performance Center, 1974.

References

Ahern, S. K. *Activation and intelligence: Pupillometric correlates of individual differences in cognitive abilities.* Unpublished doctoral dissertation, University of California, Los Angeles, 1978.

Ahern, S. K., & Beatty, J. Physiological signs of information processing vary with intelligence. *Science,* 1979, *205,* 1289–1292.

Ahern, S. K., & Beatty, J. Physiological evidence that demand for processing capacity varies with intelligence. In M. Friedman, J. P. Dos, & N. O'Connor (Eds.), *Intelligence and learning.* New York: Plenum Press, 1981.

Baddeley, A. D. A three-minute reasoning test based on grammatical transformation. *Psychonomic Science,* 1968, *10,* 341–342.

Beatty, J. Phasic not tonic pupillary responses vary with auditory vigilance performance. *Psychophysiology,* in press.

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

Beatty, J., & Wagoner, B. L. Pupillometric signs of brain activation vary with level of cognitive processing. *Science,* 1978, *199,* 1216–1218.

Bradshaw, J. L. Pupil size and problem solving. *Quarterly Journal of Experimental Psychology,* 1968, *20,* 116–122.

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

Bradshaw, J. L. Pupil size and drug state in a reaction time task. *Psychonomic Science,* 1970, *18,* 112–113.

Brodal, A. *Neurological anatomy in relation to clinical medicine* (3rd ed.). New York: Oxford University Press, 1981.

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

Goldwater, B. C. Psychological significance of pupillary movements. *Psychological Bulletin,* 1972, *77,* 340–355.

Green, D. M., & Swets, J. A. *Signal detection theory and psychophysics.* New York: Wiley, 1966.

Hakerem, G., & Sutton, S. Pupillary response at visual threshold. *Nature,* 1966, *212,* 485–486.

Hess, E. *The tell-tale eye.* New York: Van Nostrand Reinhold, 1975.

Hess, E. H., & Polt, J. H. Pupil size in relation to mental activity during simple problem solving. *Science,* 1964, *143,* 1190–1192.

Hillyard, S. A., Hink, R. F., Schwent, V. L., & Picton, T. W. Electrical signs of selective attention in the human brain. *Science,* 1973, *182,* 177–180.

Johnson, L. C., & Lubin, A. On planning psychophysiological experiments: Design, measurement, and analysis. In N. S. Greenfield & R. A. Sternbach (Eds.), *Handbook of psychophysiology.* New York: Holt, Rinehart & Winston, 1972.

Kahneman, D. *Attention and effort.* Englewood Cliffs, N.J.: Prentice-Hall, 1973.

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

Kahneman, D., & Beatty, J. Pupillary responses in a pitch-discrimination task. *Perception & Psychophysics,* 1967, *2,* 101–105.

Kahneman, D., Beatty, J., & Pollack, I. Perceptual deficit during a mental task. *Science,* 1967, *157,* 218–219.

Kahneman, D., Onuska, L., & Wolman, R. E. Effects of grouping of the pupillary response in a short-term memory task. *Quarterly Journal of Experimental Psychology,* 1968, *20,* 309–311.

Kahneman, D., Peavler, W. S., & Onuska, L. Effects

of verbalization and incentive on the pupil response to mental activity. *Canadian Journal of Psychology,* 1968, *22,* 186–196.

Kinsbourne, M., & Hicks, R. E. Functional cerebral space: A model for overflow, transfer and interference effects in human performance: A tutorial review. In J. Requin (Ed.), *Attention and performance VII.* Hillsdale, N.J.: Erlbaum, 1978.

Lang, P. J., Rice, D. G., & Sternbach, R. A. The psychophysiology of emotion. In N. S. Greenfield & R. A. Sternbach (Eds.), *Handbook of psychophysiology.* New York: Holt, Rinehart & Winston, 1972.

Lowenstein, O., & Loewenfeld, I. E. The pupil. In H. Davson (Ed.), *The eye* (Vol. 3). New York: Academic Press, 1962.

Luria, A. R. *The working brain.* New York: Basic Books, 1973.

Miller, G. A. The magical number seven, plus or minus two: Some limits of our capacity for processing information. *Psychological Review,* 1956, *63,* 81–97.

Moray, N. Where is capacity limited? A survey and a model. *Acta Psychologica,* 1967, *27,* 84–92.

Moruzzi, G. *The sleep–waking cycle. Reviews of physiology: Biochemistry and experimental pharmacology.* New York: Springer-Verlag, 1972.

Navon, D., & Gopher, D. On the economy of the human-processing system. *Psychological Review,* 1979, *86,* 214–255.

Navon, D., & Gopher, D. Task difficulty, resources, and dual-task performance. In R. S. Nickerson (Ed.), *Attention and performance VIII.* Hillsdale, N.J.: Erlbaum, 1980.

Norman, D. A., & Bobrow, D. J. On data-limited and resource-limited processes. *Cognitive Psychology,* 1975, *7,* 44–64.

Parasuraman, R. Memory load and event rate control

sensitivity decrements in sustained attention. *Science,* 1979, *205,* 924–927.

Parasuraman, R., & Davies, D. R. A taxonomic analysis of vigilance performance. In R. R. Mackie (Ed.), *Vigilance: Theory, operational performance, and physiological correlates.* New York: Plenum Press, 1977.

Payne, D. T., Parry, M. E., & Harasymiw, S. J. Percentage pupillary dilation as a measure of item difficulty. *Perception & Psychophysics,* 1968, *4,* 139–143.

Peavler, W. S. Individual differences in pupil size and performance. In M. Janisse (Ed.), *Pupillary dynamics and behavior.* New York: Plenum Press, 1974.

Posner, M. L., & Boies, S. J. Components of attention. *Psychological Review,* 1971, *76,* 391–405.

Posner, M. L., & Mitchell, R. F. Chronometric analysis of classification. *Psychological Review,* 1967, *74,* 392–409.

Ryan, J. F. *Grouping and short-term memory.* Unpublished doctoral dissertation, University of Cambridge, 1967.

Sanders, A. F. Some remarks on mental load. In N. Moray (Ed.), *Mental workload: Theory and measurement.* New York: Plenum Press, 1979.

Spielberger, C. D. *State–trait anxiety inventory.* Palo Alto, Calif.: Consulting Psychologists Press, 1968.

Wickens, C. D. Human operator workload. In N. Moray (Ed.), *Mental workload: Theory and measurement.* New York: Plenum Press, 1979.

Wickens, C. D. The structure of attentional resources. In R. S. Nickerson (Ed.), *Attention and performance VII.* Hillsdale, N.J.: Erlbaum, 1980.

Wright, P., & Kahneman, D. Evidence for alternative strategies of sentence retention. *Quarterly Journal of Experimental Psychology,* 1971, *23,* 197–213.