Motivation and Emotion, Vol. 3, No. 4, 1979

The Pupillary Response as an Indicator of Arousal and Cognition

Robert F. Stanners,[1] Michelle Coulter, Allen W. Sweet, and Philip Murphy
Oklahoma State University

Psychological research on the pupillary response since 1960 has focused on an arousal interpretation and a cognitive interpretation. Experiment 1 was an attempt to manipulate some arousal factors while controlling the cognitive demands of the task. Pupil size was cinematographically recorded while subjects who had different degrees of reported fear of snakes listened to passages describing imagined interactions with a snake in different proximities. There was also a set of control passages that made no mention of snakes but were otherwise semantically and syntactically identical to the aversive passages. The pupillary response showed no influence of the arousal manipulations, but rating and behavioral data indicated that the arousal variables had been effective. The cognitive demands of the task were clearly indicated by the pupillary response. In Experiment 2 two types of tasks were used: one that employed both arousal (incentive) and cognitive factors and another that had an arousal manipulation (threat of shock) but no explicit cognitive demands. The pupil response was recorded as well as heart rate, skin conductance, and EMG. The pupillary response showed an effect of the arousal manipulations only when cognitive demands were minimal. The results of both experiments are consistent with the view that cognitive demands take priority over arousal factors in affecting the pupillary response. Heart rate did show arousal effects that were not preempted by cognitive demands.

In the period of approximately two decades since the reintroduction of the pupillary response into psychological research (Hess & Polt, 1960), two general psychological interpretations of the response have been proposed.

[1]Address all correspondence to Robert F. Stanners, Department of Psychology, Oklahoma State University, Stillwater, Oklahoma 74074.

0146-7239/79/1200-0319$03.00/0

An early interpretation was to view the pupillary response as reflecting level of arousal or emotionality (Hess & Polt, 1960). Somewhat later (Hess & Polt, 1964; Kahneman & Beatty, 1966), it was argued that the pupillary response might reflect cognitive activity as well as or instead of arousal. A general idea of the difference between these interpretations is conveyed by the following quotation from Kahneman and Peavler (1969, p. 317): "Arousal is often constructed as an essentially automatic reaction to significant or overwhelming stimuli, whereas processing load refers to the demands imposed by activities in which *S* engages, often voluntarily." A great deal of empirical research—reviewed by Goldwater (1972), Hess (1972), and Janisse (1973)—can be interpreted equally well by either an arousal or a cognitive conceptualization. A much smaller group of studies, considered later, provides evidence that appears to weigh more heavily in favor of one or the other of the two interpretations. The major objective of the present investigation was to provide further conceptual and empirical clarification of the arousal and cognitive interpretations of the pupillary response.

The initial preference of investigators was an arousal interpretation, and this preference appears to have been the result of two types of evidence. In retrospect, however, neither of these sources of evidence provides unmixed support for the arousal interpretation. First, there was the physiological evidence. Sympathetic innervation of pupil dilation has been documented, although parasympathetic inhibition is also involved (Lowenstein & Loewenfeld, 1962). Another source of physiological information favoring an arousal interpretation was the evidence that hypothalamic destruction produced a substantial reduction in the size of the dilation response (Lowenstein & Loewenfeld, 1962). However, there is also physiological evidence that would be consistent with a cognitive influence on the pupillary response. Decorticated animals show exceptionally strong dilation responses (Lowenstein & Loewenfeld, 1962). Liberman (1965), who viewed the pupillary response as a component of the orienting response, interpreted the large and inextinguishable dilation responses in decorticates as indicating a dual control system for the pupil. According to Liberman (p. 188), "we see the physiological mechanism of the orienting pupillary reaction as follows: its executing nervous mechanism is located in subcortical structures, and its regulation is performed by the cortex."

Another apparent influence contributing to the arousal interpretation of the pupillary response was the type of variables investigated in the earlier research. The great bulk of the earlier experimental studies of the pupil in the modern psychological literature (see the reviews by Goldwater, 1972; Hess, 1972; Janisse, 1973) were concerned with variables that could be roughly categorized as interest or as positive and negative affective states. There was also a theoretical view, with some empirical evidence, which appeared to provide strong support for an arousal interpretation. Hess (1965) proposed a biphasic hypothesis of pupillary activity in which positive affect was indicated by pupil dilation and negative affect was signaled by pupillary constriction. This type of pupillary action went very well with an arousal interpretation since affect states have meaningful positive and negative values in the subjective sense. The biphasic view of pupillary action has, however, been so extensively criticized that rehabilitation of the hypothesis appears doubtful. A detailed presentation of the critical evidence is available in Janisse (1973). Some of the lines of criticism are (a) the types of pictorial materials used by Hess may have produced constriction due to the light reflex (Loewenfeld, 1966) and also have led to conflicting results (Woodmansee, 1966); (b) visually presented words, for which light values can be better controlled than for pictures, have produced dilation for both positive and negative semantic connotations (Vacchiano, Strauss, Ryan, & Hochman, 1968); (c) constriction has not been found for affectively negative mental images (Paivio & Simpson, 1966). There is, in addition, a plausible cognitive account of occasional biphasic activity. Affectively positive stimuli could simply stimulate a higher level of thinking; affectively negative stimuli could reduce the level of cognitive activity below the baseline level that prevailed prior to stimulus presentation, thereby producing apparent constriction.

The cognitive interpretation of the pupillary response in the modern psychological literature made its appearance in an experiment by Hess and Polt (1964) in which subjects were given multiplication problems of varying levels of difficulty while pupil size was recorded. There was an increase in pupil size after the problem was presented, and the size of the increase was related to the apparent difficulty level of the problem. These results can be readily interpreted as indicating that pupil size reflects the level of cognitive effort, but can just as plausibly be interpreted as an arousal effect associated with the perceived difficulty of the problem. A procedure with somewhat better prospects of differentiating the arousal and cognitive interpretations was introduced by Kahneman and Beatty (1966), and the basic results have since been replicated numerous times (Goldwater, 1972; Janisse, 1973). Subjects were given a short-term memory task in which digits or words were presented at a 1-second interval, followed by a 2-second pause, which was in turn followed by a recall signal. The major results, from the point of view of differentiating the arousal and cognitive interpretations, were the "loading" and "unloading" functions. The size of the pupil increased in a second-by-second fashion with each successive memory item and then decreased in a similar fashion as each of the items was reported. The momentary variation in pupil size in conjunction with task demands appears to be characteristic of a cognitive process rather than of a state of arousal.

Some studies have attempted to evaluate the arousal and cognitive interpretations by means of experiments in which both types of factors are covaried. Kahneman, Peavler, and Onuska (1968) engaged subjects in a digit transformation and short-term memory task (a value is added to each of a series of digits and the transformed digits are reported back) in which difficulty (add 0 versus add 1) and incentive (2 cents or 10 cents for each correct response) were manipulated. A small effect of incentive appeared in only the easier task. The interpretation of the incentive effect was that greater incentive produced greater cognitive effort. An arousal interpretation did not appear attractive since there was no incentive effect in the more difficult task. Presumably, the level of cognitive effort was already sufficiently high to preclude an additional (cognitive) effect due to incentive. Kahneman and Peavler (1969) manipulated incentive in a task involving the learning of digit–noun paired associates. The digits and nouns were alternately presented at a 3-second interval, and odd or even digits signaled differential incentive for a given pair. Larger dilations were found on high incentive pairs but only for presentation of the nouns (the response term) not of the digits. A cognitive interpretation based on effort was adopted on the basis that a pupillary response due to arousal should have appeared when the digits were presented.

Negative incentive and difficulty level were varied in an experiment by Polt (1970). Multiplication problems of different levels of judged difficulty were presented to separate groups of subjects who either were given no negative incentive or were told in advance that they would be shocked for an incorrect answer. There were larger dilations associated with the negative incentive, but they occurred only after the problem had been presented and not during a control period just prior to presentation of the problem. Polt viewed the lack of an effect during the control period as favoring a cognitive interpretation based on additional effort induced by the threat of shock.

The results of the incentive studies can be accommodated by a strictly cognitive interpretation, but there are some results that do not appear to have the characteristics of a cognitive process. Nunnally, Knott, Duchnowski, and Parker (1967) told subjects that the digits 1–5 would be presented, and that a loud gunshot would occur during the digit 3. Pupil size increased during digits 1–3 and decreased during digits 4–5. A cognitive interpretation might argue that the subject exerts a higher level of mental effort during the digits 1–3, but such a hypothetical cognitive process not related to task demands seems gratuitous. This type of result appears to have the anticipation character of an arousal process. Simpson and Molloy (1971) categorized subjects on the basis of a questionnaire designed to measure anxiety in front of an audience. The subjects performed a digit transformation and short-term memory task that involved a period of listening to the digits, a 3-second pause, and than a report to two experimenters. There was no difference between high and low audience-anxiety groups during the listening period, but a reliable difference in favor of the high group did appear during the pause period. A cognitive interpretation based on differential effort does not appear attractive because the difference in pupil size occurred during the pause period and not during the listening period. An anticipation-arousal interpretation seems more plausible. Stanners and Headley (1972) also employed an anticipation manipulation that produced results indicating an arousal effect. Subjects were given one of four instructions prior to each new trial indicating what kinds of manipulations they would be required to perform on the material; the instruction was followed by a 4-second pause, and then the material was presented. Pupil size for the condition that involved the most difficult task manipulations remained at the postinstruction level throughout the pause period. Pupil size for conditions involving less difficult task manipulations drifted downward to a level corresponding to relative task difficulty. Since the differences in size occurred prior to the actual presentation of the materials, a cognitive interpretation did not appear plausible.

EXPERIMENT 1

This review of efforts to provide a psychological interpretation of the pupillary response indicates that there may be both arousal and cognitive effects. In general, a close temporal linkage between pupillary behavior and immediate task demands would support a cognitive effect. On the other hand, pupillary changes that are not associated with, or occur prior to, task demands would tend to indicate an arousal effect. It seems clear that any attempt to isolate one of the effects must control for the other one. In Experiment 1, three arousal variables were manipulated while attempting to hold constant the cognitive demands of the situation. One of the arousal variables was a trait type of variable, snake phobia. Its choice was based on the fact that it is possible to find substantial differences in a normal population and, also, it can be indexed in three separate ways: physiologically, through ratings of subjective anxiety, and by a behavioral test. A state type of arousal variable was the level of induction of fear associated with snakes, which was manipulated by the construction of short passages that described imagined interactions between a subject and a snake in different degrees of proximity to one another. A third arousal manipulation came about as a result of the control condition for cognitive effort. A parallel set of passages was constructed that substituted another object for the reference to a snake but were otherwise semantically and syntactically identical.

Method

Materials. The materials consisted of a series of statements describing a situation (scene) that the subject was to imagine. The scenes comprised three levels of aversiveness, which depended on the described proximity of a snake to the subject. For example, one of the low-aversive scenes was, "Imagine sitting in a car, and you see a snake go across the road." A medium-level aversive scene was, "Imagine that you are in a pet store, and you see a snake in a wire cage." An example of a high-level aversive scene was, "Imagine lying on a divan with a large snake curled up on your stomach." There were three exemplars for each level of scene. Associated with the aversive scenes was a set of nine neutral scenes that retained the syntactic and some of the sematic characteristics of the aversive counterparts. The neutral scenes were constructed by changing one or two words in such a way as to eliminate the reference to a snake and yet to produce a plausible imaginal scene. The neutral counterparts of the aversive examples were constructed as follows: substitute "dog" and "kitten" for "snake" in the first two examples. Substitute "teddy bear cuddled" for "large snake curled" in the high-level aversive example.

The scenes were used for two different purposes. One usage was as an experimental manipulation of fear. The other use to which the scenes were put was a selection device for subjects. For this latter purpose the scenes were reworded into the same form as statements on the Wolpe Discomfort Index (Wolpe and Lazarus, 1966). Examples of the reworded scenes are: "Sitting in a car and seeing a snake cross the road" and "Seeing a snake in a wire cage in a pet store."

Subjects. The subject pool was composed of the female students in an introductory psychology class in which participation in research was one way to gain an extra point bonus. Of the female students in the class, 57 were willing to participate as subjects and were given the Wolpe Discomfort Index, which included the nine reworded scenes. Each scene was rated on a scale from 0 to 4 on the basis of how intense a level of fear would be produced by the event described in the statement. Eight subjects in each of three groups were selected by the criterion of the average rating on the nine reworded aversive scenes. The low-fear group had an average rating of 10.5 for the nine items, the average rating for the medium group was 16.9, and the average for the high-fear group was 27.9. The total possible range of scores on the nine statements was 0–36.

Equipment. The system for recording pupil size was a Hess-type pupillometer (Hess, 1965). Its essential features were as follows: The viewing system was a rectangular plywood box, measuring 58 cm X 58 cm X 123 cm, fitted at one end of the long axis with a translucent screen. The other end was completely closed except for an eyepiece and an adjustable chin rest. The interior of the box was painted flat black and fitted with a half morror mounted at 45° to the vertical plane of the subject's line of sight in such a way that the subject's right eye was reflected into the camera lens system. The camera, a Beaulieu R16ES, was mounted at the side of the pupillometer box with a 30-mm extension tube, which protruded approximately 1 cm into the box through a close-fitting aperture. Automatic control of the camera was provided by a sound-operated relay activated by a signal on one channel of the tape recorder (Uher Royal Deluxe) used to present the materials. Another sound-operated relay activated a pinhole light source below the eyepiece of the pupillometer, which was used to mark the film for various trial events. Illumination in the pupillometer at the eye piece was 18.3 mL and was provided by two sources, a 200-watt incandescent bulb mounted behind the translucent screen and five miniature 12-V bulbs mounted in the eyepiece. Filming took place at two frames per second, and Kodak High Speed Infrared Type 2481 film was used. Infrared film permitted the use of subjects with any iris color.

Procedure. When the subject arrived for the experimental session, she was shown the apparatus and told in a general way about its use. After the adjustments of the chin rest and camera focus, headphones were placed on the subject, and instructions were played from the tape recorder. All instructions and experimental material were presented auditorily by tape. The instructions indicated that the subject should fixate a small cross on the translucent screen during the trial, that she should inhibit eyeblinks, and that she should attempt to visualize the scene as vividly as possible. The sequence of events on a trial was as follows: First, the subject heard the word *ready,* at which time she was supposed to fixate the cross and try not to blink. Four seconds later the visualization scene was begun, which lasted 4–6 seconds. On the 10th second after the ready signal, the word *visualize* was presented, and on the 15th second, the word *report.* When the report signal was given, the subject stated two digits, which were values on a 7-point scale indicating (a) how vividly the scene had been visualized and (b) how anxiety-provoking the scene had been. On the 25th second after the ready signal, the instruction *rest* was given, followed by a 35-second break. Filming began with the ready signal and terminated with the report signal.

The ordering of the scenes was random, with the restriction that no more than three aversive or neutral scenes occur in a row. A second, "mirror image" order of the scenes was used in which each aversive scene was replaced by its neutral counterpart and vice versa. Each order was given to four subjects within each of the fear groups. Presentation of the 18 experimental scenes was preceded by 4 practice scenes, which were ostensibly neutral and easy to visualize.

After all the scenes had been presented, the subject was taken to an adjacent room and given the behavioral avoidance test. The room was bare except for a table and a glass cage with a wire mesh top in which there was a bull snake approximately 1.4 m in length. Starting at the doorway of the room, 3 m from the cage, the subject was told to go into the room and observe the snake. After the subject stopped approaching the cage, she was told to put on a leather glove, open a trap door in the top of the cage, reach in the opening, and touch the snake. If the subject asked the experimenter if he really wanted her to touch the snake, the standard reply was, "I would like for you to." After the behavior avoidance test, the subject gave a verbal pledge not to discuss the experiment with any classmates and was then dismissed.

Data Scoring. Each frame of film was projected by a Xerox Model 2240 40X microfilm reader, and pupil diameter was measured to the nearest mm with a transparent ruler. This procedure produced accuracy to the nearest .1 mm of actual pupil size. Five percent of the total of 12,960 frames were not scorable due to head movements, eyeblinks, or poor focus. The measurement for each frame was expressed as a deviation from the average of the 8 baseline frames that preceded the presentation of a scene. An individual score was the average of the deviations for a 1-second interval (2 frames) of the three exemplar scenes at a given level of aversiveness or the control counterparts for a total of six deviations.

Results

The overall analysis of the pupil data was a five-factor analysis of variance with the following factors: Time (5 seconds of filming), Period (listening and visualization), Group (low, medium, and high fear), Type of Scene (aversive or neutral), and Aversiveness Level of the Scene (low, medium, and high). An overall arousal effect of the scenes, given the use of matched neutral scenes, would be expected to appear in a main effect of Type of Scene. A graded arousal effect, dependent on the aversiveness level of the scene, should appear in a Type of Scene by Aversiveness Level interaction effect. The first two frames of film in the listening period were discarded to make the filming time compatible with the visualization period.

Since four of the five factors were within subjects and the Time factor was by definition sequential, the conservative Geisser and Greenhouse (1958) correction was performed on the within-subjects tests to compensate for a highly probable violation of the assumption concerning a symmetrical variance–covariance matrix.[2] The .05 level was adopted as the minimum for

[2]The effect of the Geisser and Greenhouse (1958) correction procedure was to reduce the degrees of freedom for testing within-subjects effects to 1/21. The actual degress of freedom from the design are given in the text.

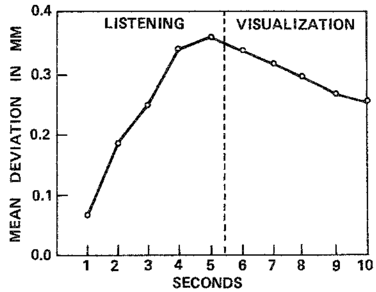

Fig. 1. Average change in pupil size as a function of Period and Time.

considering an effect significant. The significant main effects in the overall analysis were Time ($F(4/84) = 15.35$, $p < .001$) and Period ($F(1/21) = 5.09$, $p < .05$). The significant interaction effects were Time by Period ($F(4/84) = 75.92$, $p < .001$) and Period by Aversiveness Level of the Scene ($F(2/24) = 5.32$, $p < .05$). The form of the main effects and the Time by Period interaction can readily be seen in Figure 1. The Period by Aversiveness Level interaction was produced by a slightly different ordering of average pupil size for the three levels of scenes during the listening and visualization periods, as well as a change in the size of the mean differences. During listening, the high, medium, and low level scenes produced mean pupil dilation changes of .32, .22, and .23, respectively. During visualization these values were .32, .26, and .25.

The significant interactions prompted separate analyses for the listening and visualization periods. In the listening period the only significant effect was Time ($F(4/84) = 59.45$, $p < .001$). In the visualization period Time was likewise the only significant effect ($F(4/84) = 7.23$, $p < .025$). In neither period did Aversiveness Level, which was involved in an interaction with Period in the overall analysis, produce a significant effect.

After the visualization period for each scene, the subjects were required to give a rating on a 7-point scale reflecting (a) the vividness or intensity of the image and (b) the subjective anxiety produced by the scene. Scores were produced by averaging the ratings for the three exemplars at a given level of aversiveness. For each type of rating a three-factor analysis of variance was performed with the following factors: Group (low, medium, and high fear), Type of Scene (aversive or neutral), and Aversiveness Level of the Scene (low, medium, and high). The Geisser and Greenhouse (1958) adjustment was applied in the tests of the within-subjects terms. The analysis of the ratings of image intensity produced no significant effects, so any differences in the difficulty of imagining the various scenes were unrelated to the experimental variables. All the main effects of the ratings of subjective anxiety were significant. For the fear groups, $F(2/21) = 4.16$, $p < .05$; for type of scene, $F(1/21) = 78.08$, $p < .001$; and for level of scene, $F(2/42) = 34.52$, $p < .001$. The significant interaction effects were Group by Type of Scene ($F(2/21) = 4.76$, $p < .05$) and Aversiveness Level by Type of Scene ($F(4/42) = 34.81$, $p < .001$). For both of the interaction effects the pattern of means in Table I indicates that the group and level of scene manipulations were producing effects for the aversive scenes but not for the neutral scenes. Separate analyses of variance were performed for each type of scene. There were no significant effects for the neutral scenes, and the differences among means in Table I are small and not consistent with the ordering of the variables. For the fear scenes there were two significant effects: Group ($F(2/21) = 8.36$, $p < .005$) and Aversiveness Level of Scene ($F(2/42) = 62.55$, $p < .001$). Table I indicates that the ordering of the means for these effects was consistent with the levels of the variables.

There was one further set of data that was produced by the behavioral avoidance test. Subjects in each of the fear groups were classified on the basis of whether they did or did not actually touch the snake. None of the subjects in the high fear group touched the snake, three subjects in the medium group did, and six subjects in the low fear group did. A chi-square analysis of the frequency table produced by fear group and touch or no touch yielded a significant chi-square of 9.6 ($df = 2$, $p < .01$).

Discussion

The evidence from the pupillary response in favor of an arousal effect was negligible. There were three arousal manipulations: fear group, aversive versus neutral scenes, and aversiveness level of the scenes. None of the main effects or interactions involving the arousal variables was significant except for the Aversiveness Level by Period interaction. This effect did not hold up when separate analyses were performed on the listening and visualization periods.

Table I. Mean Anxiety Ratings

Level of scene	Fear group: Low	Medium	High
Aversive scenes			
Low	1.96	2.83	3.33
Medium	2.71	4.00	4.58
High	4.50	6.67	6.50
Neutral scenes			
Low	2.16	1.71	1.29
Medium	1.92	2.00	2.25
High	1.58	2.25	1.91

The results of the analyses of the anxiety ratings were exactly what would be expected from an arousal effect. There was a difference between aversive and neutral scenes, a graded effect of levels of aversive scenes, and a fear group effect. The rating results associated with the scenes could, however, also have been produced by the subjects' semantic interpretation of the scenes. That is, the subjects could have based their anxiety ratings on a comparison of the language used in describing the scenes rather than on the actual fear that the scenes elicited. The effect of fear group on anxiety ratings could not be explained by a semantic comparison of the statements but could nonetheless have been produced by another cognitive effect. The effect of fear group on the rating results could have been produced by subjects basing their ratings on how the individual subject *would* feel in the actual situation described in the scene rather than on the anxiety actually elicited by the description. It is possible, of course, that the rating results were due to fear; the cognitive interpretation is simply an alternative. The cognitive explanation would not seem to be appropriate for the results of the behavioral avoidance test since the subjects were reacting to an actual situation rather than to a description. The behavioral test indicates that subjects in the different fear groups did have different levels of fear insofar as approach behavior can be taken as an indicator.

Whatever arousal effects may have been produced by the experimental variables were not picked up by the pupillary response. On the other hand, the cognitive effects on the pupil were very substantial. Pupil size increased in a very regular fashion during the listening period, as would be expected from previous research involving sentence listening (Wright & Kahneman, 1971; Stanners, Headley, & Clark, 1972). If an arousal state had been produced by any of the arousal variables, the effects on the pupillary response may have been overridden by the cognitive demands of the task. There may be a hierarchy of control for the pupillary response such that cognitive demands take first priority and arousal effects show up only if cognitive demands are minimal.

EXPERIMENT 2

Another approach to the problem of disentangling cognitive and arousal influences on the pupillary response would be to employ other measures of autonomic activity in conjunction with the measurement of pupil activity. An advantage of the use of multiple measurements would be in providing other objective criteria of whether or not a given arousal manipulation had been successful. It could be quite informative if an arousal manipulation produced no effect on the pupillary response but did produce an effect on skin conductance or heart rate. The measurements used in Experiment 2 were the pupillary response, heart rate, skin conductance, and EMG. The experimental task involved digit transformation and short-term memory at different difficulty levels in conjunction with both positive and negative incentives. In addition, there was an arousal manipulation (threat of shock) in the absence of any explicit cognitive demands. If there is a priority of cognitive control of the pupillary response, then such a task would be expected to allow an arousal effect to appear. Another objective of Experiment 2 was to investigate the possibility of idiosyncratic reactions to stress. Several previous studies have indicated a difference in physiological response patterns between persons with internal somatic symptoms (for example, stomach distress, headaches, colitis) and those with external types of complaints (skin disorders, arthritis). Fisher and Cleveland (1960) found that under stress arthritics showed a greater number of GSR responses and a lower heart rate than ulcer patients. Moos and Engel (1962) reported higher levels of muscle potential and GSR in subjects with rheumatoid arthritis as compared to hypertensives. Fisher (1970) summarized this line of research in the form of a principle of individual response stereotypy, which states that physiological systems that match the locus of bodily symptoms will be more highly reactive under stress. An implication of this principle of concern in the present experiment is that the effect of a stressor might be more precisely ascertained by inclusion of a subject variable that differentiates among internal symptoms, external symptoms, and neither type of symptom.

Method

Subjects. Thirty male subjects were selected from undergraduates at Oklahoma State University on the basis of their willingness to participate in a procedure that might involve electrical shock and performance on a modified version of the Cornell Medical Index. Ten subjects were selected for each of three groups: an external symptom group (for example, sensitive skin, rashes, acne), an internal symptom group (for example, upset stomach, headaches, colitis), and a neutral group showing neither type of symptom. Subjects in the symptom groups gave one or more affirmative responses to items of the appropriate category; those in the external group indicated an average of 2.6 symptoms and those in the internal group an average of 1.4 symptoms.

Equipment. The pupillometer was the same as the one used in Experiment 1. The recording system for skin conductance, EMG, and heart rate was a Physiograph Six (Narco Bio-Systems, Inc.). Heart rate was monitored by a silver disk electrode placed on the upper left area of the subject's chest with a ground electrode on the lower right side of the chest. Electromyographic activity was monitored by two silver disk electrodes placed above the forearm flexors on the right arm approximately 10 cm apart. Skin conductance was recorded from two silver-silver chloride cup electrodes placed on the palm and ventral wrist of the subject's left hand. Skin surfaces for all electrode placements were cleansed with alcohol, lightly scrubbed with fine sandpaper, and coated with Redux Creme.

Procedure. The subject was seated before the pupillometer while the recording electrodes were attached. After a 10-minute interval, during which equipment was being adjusted, a tape-recorded set of instructions was played through headphones in which the subject was told that some bodily responses would be recorded while he performed some problems. The type of problem was described to the subject as one in which a digit would be presented that was to be added to each of four subsequently presented digits. When the signal *respond* occurred, the four transformed digits were to be reported to the experimenter. The subject was then told that he would be given three practice problems. The sequence of events on a problem was as follows: The subject positioned his head in the eyepiece of the pupillometer with his chin in the chin rest. The word *attention* was then presented, which was followed by a 3-second interval and then the word *ready*. The ready signal corresponds to the 1st second in Figures 2–4. The instruction to add 0, 1, or 3 was presented on seconds 3–4. The four digits

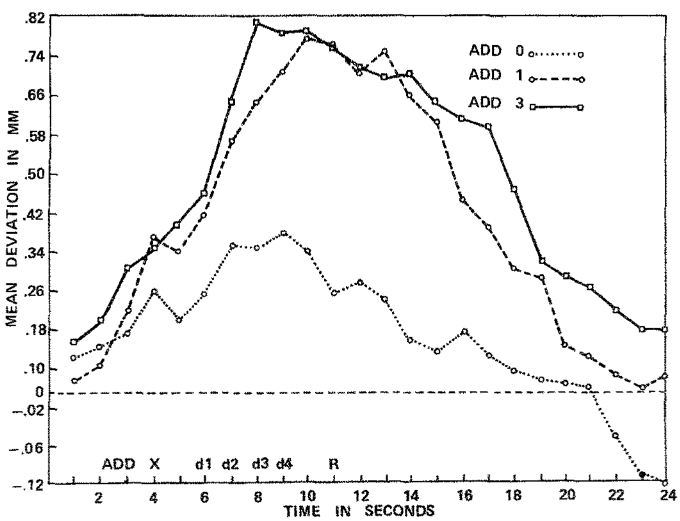

Fig. 2. Average change in pupil size as a function of Difficulty Level and Time.

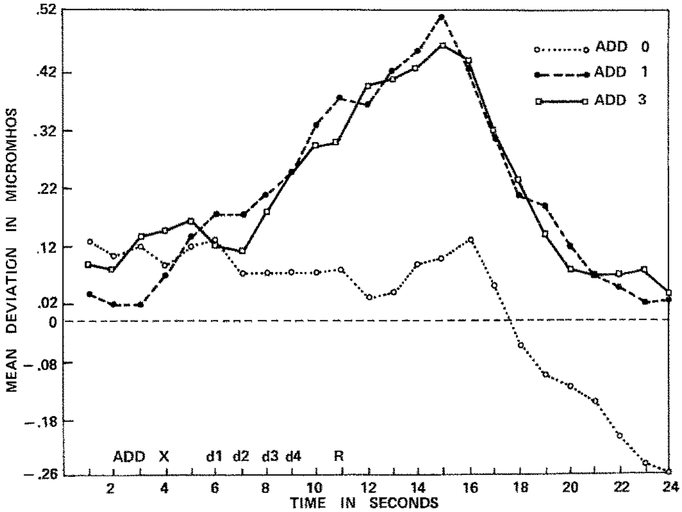

Fig. 3. Average change in skin conductance as a function of Difficulty Level and Time.

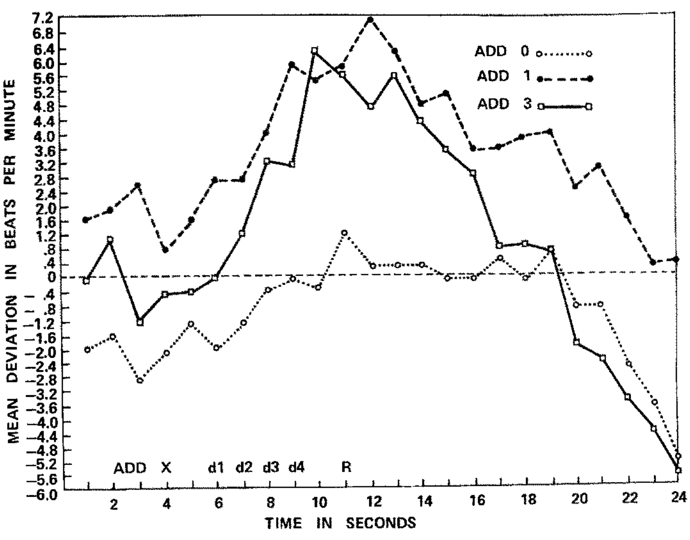

Fig. 4. Average change in heart rate as a function of Difficulty Level and Time.

were presented at 1-second intervals on seconds 6–9, and on the 11th second, the instruction *respond* was given. The instruction *relax* was given on the 24th second, which signaled a 10-second break, during which the subject could close his eyes but remained in the chin rest. Two further problems followed, with the same format but different digits. The subject was then told to ask any questions he might have about the task and informed that the practice problems were over. Another set of three problems with the same format was presented and followed by further instructions. These instructions indicated that on the next set of problems the subject would be rewarded with 5 cents for each problem performed correctly, and the subject was paid following the problems. Further instructions now informed the subject that he would be punished for incorrect responses on the following set of problems by a mild electrical shock to the neck. Incorrect responses were defined as those that fell outside an undisclosed range of error, and the subjects were told that about 50% of the people who attempt the problems get them all within the range. Electrodes were then attached to the subject's neck, and another set of problems was presented. No shocks were actually given, the implication being that all responses were within the permissible range. The value used in the add instruction, which occurred on seconds 3–4 during a given problem, followed a specific sequence during the practice session and the other problem sessions. The sequence was, add 0, add 1, and add 3.

In the final phase of the procedure, the subject was informed that some measurements of reactions to shock were needed from subjects who had not received any shocks during the last problem session. The shock was to be given on the basis of a coin flip, on which the subject was to call "heads" or "tails," but no information about the outcome of the flip would be given until after the complete procedure. The (countdown) procedure consisted of the auditory presentation of the numbers 1–10, with the shock coming on the number 10 if the subject had "lost" on the coin flip. After the subject positioned his head in the pupillometer, there was a 3-second interval, and then the numbers 1–10 were presented at a 2-second interval. Recording was continued for 20 seconds beyond the number 10, at which time the subject was debriefed as to the purpose of experiment, and his cooperation in not discussing it was solicited.

Data Scoring. The scoring of the pupillary data was essentially the same as in Experiment 1. Each frame measurement (taken to the nearest .1 mm) was expressed as a deviation by subtracting the average of the six baseline frames that were recorded between the attention signal and the ready signal. An individual score was the average of the two deviations for a 1-second interval under a given combination of conditions. A similar type of scoring system was used for skin conductance and heart rate, except that they were sampled only once per second. Skin conductance was measured in micromhos on each of the 1-second marks of the inkwriter record. The average of the three values between "attention" and "ready" was subtracted from each value taken subsequent to the ready signal. The interval between R waves for the heart was converted into beats per minute and expressed as a deviation by subtracting the average of the three values taken on each trial between the attention and ready signals. For EMG, an individual score was the value in microvolts from a zero baseline integrated over a 1-second interval. The scoring procedures for the countdown were the same as for the digit transformation task, except that data values were averaged over a 2-second interval rather than a 1-second interval.

Results

The first set of analyses was concerned with performance during the digit transformation task. For each of the four measures, a four-factor analysis of variance was performed that had the following structure: Symptom Group (internal, external, and neutral), Difficulty Level (add 0, add 1, add 3), Incentive (no incentive, reward, threat of shock), and Time (24 seconds). The Geisser and Greenhouse (1958) correction was again applied to the within-subjects effects, and the .05 level was taken as the minimum for an effect to be considered significant.[3]

For the pupil response there were two significant main effects: Difficulty Level ($F(2/54) = 25.28$, $p < .001$) and Time ($F(23/621) = 97.67$, $p < .001$). The interaction of difficulty and time was also significant ($F(46/1242) = 15.57$, $p < .001$). The nature of the interaction is apparent in Figure 2; the time function for the add 0 condition is very different in shape from those of the other difficulty levels. It is also apparent that the main effect of difficulty is attributable to the difference between the add 3 and add 1 conditions on the one hand and the add 0 condition on the other. Neither of the direct arousal manipulations (symptom group or incentive) produced any significant effects during the digit transformation task, although there is an effect that could be interpreted as anticipation or preparation. Pupil size in the add 1 and add 3 conditions shows an increase that occurs prior to the actual presentation of the first digit at the 6th second. The simple effect of difficulty at the 5th second was significant ($F(2/54) = 9.31$, $p < .01$) but was not significant at seconds prior to the 5th.

The significant main effects for skin conductance were Difficulty ($F(2/54) = 4.99$, $p < .05$) and Time ($F(23/621) = 12.03$, $p < .005$). These

[3]The effect of the Geisser and Greenhouse correction was to reduce the degrees of freedom for testing within-subjects effects to 1/27. The actual degrees of freedom from the design are again given in the text.

factors also were involved in a significant interaction effect ($F(46/1242) = 4.73$, $p < .05$). The form of the interaction is depicted in Figure 3, in which the time functions for the add 1 and add 3 conditions are indistinguishable from each other and are both very different from the function for add 0. All of these results are very similar to those for the pupillary response. However, the apparent anticipation effect found for the pupil at the 5th second was not found for skin conductance; the first significant simple effect of difficulty did not occur until the 10th second. The longer latency of the GSR would not appear to be sufficient to account for the 5-second delay.

The analysis of variance for the heart rate data produced three main effects: Difficulty ($F(2/54) = 8.29$, $p < .01$), Time ($F(23/621) = 15.15$, $p < .001$), and Incentive ($F(2/54) = 6.54$, $p < .025$). None of the other effects were significant. The order of the overall means (difference scores averaged over 24 seconds) for the incentive manipulation was threat (2.4), reward (1.35), and task alone (−.5). Tukey HSD tests indicated a significant difference between task alone and reward and between task alone and threat but not between reward and threat. The overall means for the difficulty effect were add 0 (−1.0), add 1 (3.4), and add 3 (1.0). Multiple comparisons by the Tukey HSD test indicated a significant difference between add 0 and add 1 and between add 1 and add 3. A sketch of the time curves for difficulty is presented in Fig. 4, because the lower overall mean for add 3 as compared to add 1 is somewhat misleading. The data points for add 1 and add 3 are fairly close from the presentation of the digits (6th second) until the 16th second. The somewhat lower early slope and, particularly, the more rapid drop-off for the add 3 Time function account for most of the overall mean difference.

Only one significant effect occurred for the EMG data during the digit transformation task, a main effect of symptom group ($F(2/27) = 4.68$, $p < .025$). The form of the effect as assessed by the Tukey HSD test was that the average EMG for the external symptom group (47.26) was significantly higher than that for the internals (15.06) and also significantly higher than the mean for neutrals (18.97). The latter groups were not significantly different from each other.

The next set of analyses was concerned with the countdown procedure, in which the subject simply listened to the numbers 1 through 10 presented at a 2-second interval with an ostensible .5 probability of shock occurring on the number 10. For each of the response measures, an analysis of variance was done that has the following factors: Symptom Group (internal, external, neutral) and Time (20 levels). Time was the only significant effect for the pupil response ($F(19/513) = 28.41$, $p < .001$). Pupil size began slightly above the baseline at the 1st count (.174), dropped to virtually baseline at the 5th count (.058), and reached a maximum value at the 10th count (.627), the time of the threatened shock. Pupil size then rapidly decreased until, by 6 seconds after the 10th count, it was back at baseline and stayed there. Skin conductance also showed only a significant effect of time ($F(19/513) = 14.05$, $p < .005$). Skin conductance was well above baseline at the 1st count (.486), dropped to .352 by the 5th count, and reached a maximum 2 seconds after the 10th count (.832). It decreased monotonically from this point until, at 10 seconds after the maximum, it was very close to the baseline. Neither heart rate nor EMG produced any significant effects during the countdown procedure.

Discussion

As in Experiment 1, an arousal effect would be indicated by responses that are not linked to immediate task demands. On the other hand, a close temporal linkage between immediate task demands and responding will be taken as an indication of a cognitive effect.

The incentive manipulation did not produce any detectable effect on the pupillary response. The clearest indication of an arousal effect would have been an increase in pupil size for the positive or negative incentive conditions during portions of the trial prior to the actual introduction of the task materials. Skin conductance also failed to show any influence of the incentive manipulation. Heart rate did show an incentive effect that is compatible with an arousal interpretation. Differences in heart rate associated with the incentive conditions were present from the beginning of the trial and remained essentially consistent throughout the trial.

The difficulty manipulation produced an effect on the pupillary response, which seems most plausibly interpreted as arousal. The time curves for the add 1 and add 3 conditions became separated from that for the add 0 condition prior to (5th second) the presentation of the materials, which suggests some type of anticipation or preparatory effect similar to that found by Stanners and Headley (1972). The apparent anticipation effect did not occur for skin conductance. The difficulty manipulation produced an increase in heart rate, which occurred at the beginning of the trial, even prior to the add instruction. Since the heart rate differences occurred prior to any task demands, they are indicative of an arousal effect. The difficulty conditions were consistently ordered during the experiment. The fact that heart rate changes occurred even prior to the add instruction would seem to indicate that this response was reflecting the predictability of the difficulty conditions.

Both the pupillary response and skin conductance show what appears to be an arousal effect during the countdown procedure. There were no explicit task demands during this procedure, although it seems very likely that the subjects were listening to and interpreting the counts. Two features of the results would, however, militate against any purely cognitive interpretation of the time curves for the pupil and skin conductance. One is the fact that the curves increase throughout the count, although there was no requirement to store any information as in the digit transformation task. Simply listening to the count might be expected to produce some pupillary and skin conductance activity but should not produce a progressive increase in such activity. The other feature of the results is that the level attained by the pupillary response and skin conductance is much higher than would be predicted from cognitive activity alone. Even though task demands were minimal, both measures attained as high a level as the highest produced during the digit transformation task.

Heart rate, which did show an arousal effect for the Incentive and Difficulty manipulations, did not respond to the countdown procedure. The answer to this apparent paradox may be that the heart indicates a specific type of arousal reaction. As Elliott (1974) has argued, increased heart rate may indicate preparation for engaging in some type of activity, mental or physical. Since no activity was demanded during the countdown, preparation did not take place, and heart rate remained close to the baseline level.

Effects readily interpretable as cognitive were observed for both the pupillary response and skin conductance. There was a time effect during the digit transformation task indicating regular increases in the level of the responses with an increase in the amount of information that had to be stored. There was also an interaction of time with difficulty, in which the time curves for the add 1 and add 3 conditions diverged from that of the add 0 condition. The divergence indicates a differential response to immediate task demands (presentation of the materials). There was no comparable interaction of Time and Incentive that would have indicated different levels of cognitive effort for the different Incentive conditions. The reason for the lack of a cognitive effect of Incentive may be attributable to the task simply not being difficult enough. Kahneman (1973, chap. 2) has argued that the demands of the task place a limit on the amount of cognitive effort that can be voluntarily induced. The level of performance in the digit transformation task, as indicated by an overall error rate of only .01, would tend to support the idea that the subjects were working as hard as necessary.

The only effect for heart rate that seems interpretable as cognitive was the main effect of Time, in which the size of the heart rate change was correlated with the amount of information that had to be stored.

The Difficulty manipulation produced some other findings that suggest that there may have been some habituation occurring during the digit transformation trials. Most prominently, there is a higher mean for the add 1 condition as compared to the add 3 condition for heart rate, when the data are averaged over all time points. There is also the lack of separation of the add 1 and add 3 conditions for the pupil response and skin conductance, as has been found previously (Kahneman, Tursky, Shapiro, & Crider, 1969). Despite keeping the sessions short within the incentive conditions and using a running baseline for the pupil response, skin conductance, and heart rate, it appears that some habituation did occur.

The EMG measure produced results in sharp contrast to those of the other three measures, in showing no arousal or cognitive effects. The contrast is further sharpened by being the only measure to show an effect of symptom group. The nature of this effect was as would be predicted by Fisher's (1970) principle of response stereotype; the external symptom group had a much higher overall level of EMG activity during the digit transformation tasks than did either the internal or neutral groups.

GENERAL DISCUSSION

The results of Experiments 1 and 2 considered jointly tend to support a hierarchical control system for the pupil response, as has been suggested by Liberman (1965), and also place some constraints on the conditions under which an arousal or a cognitive effect can be indicated by the pupil. In Experiment 1, three different arousal manipulations failed to produce any effect while task demands were held constant. Cognitive effects, however, were clearly indicated by the pupil response. One interpretation of the results is that substantial cortical involvement, as required by language interpretation, "preempts" pupillary control to the exclusion of arousal factors. The arousal effects that were observed in Experiment 2 for the pupillary response are consistent with this interpretation. There was an effect during the countdown when there were no explicit task demands except listening to the count. There was also an arousal effect on the digit transformation trials that occurred prior to the presentation of the materials. However, after the presentation of the task materials, the pupillary response showed only cognitive effects. Heart rate, on the other hand, showed arousal effects of the Difficulty and Incentive manipulations, but no cognitive effects of these variables.

The following conclusions are suggested by the results of Experiments 1 and 2: The pupil response will show an arousal effect only when the cognitive demands of the situation are minimal. The control system is such that if the situation requires a substantial level of cognitive activity, only this will be indicated by the pupillary response. Some indication as to the level of cognitive activity required to preempt the indication of arousal is provided by the tasks used in Experiments 1 and 2 but can be more precisely ascertained by further experimentation. Heart rate, in contrast to the pupillary response, will show an arousal effect that is not preempted by cognitive demands, but only under conditions where the situation requires preparation for some type of mental or physical activity.

REFERENCES

Elliot, R. The motivational significance of heart rate. In P. Obrist, M. H. Black, J. Brener, & L. DiCara (Eds.), *Cardiovascular psychophysiology.* Chicago: Aldine, 1974.

Fisher, S. *Body experience in fantasy and behavior.* New York: Appleton-Century-Crofts, 1970.

Fisher, S., & Cleveland, S. E. A comparison of psychological and physiological reactivity in ulcer and rheumatoid arthritis groups. *Psychosomatic Medicine,* 1960, *22,* 290-293.

Geisser, S., & Greenhouse, S. W. An extension of Box's results of the use of the *F* distribution in multivariate analysis. *Annals of Mathematical Statistics,* 1958, *29,* 885-891.

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

Hess, E. H. Attitude and pupil size. *Scientific American,* 1965, *212,* 46-54.

Hess, E. H. Pupillometrics: A method of studying mental, emotional, and sensory processes. In N. S. Greenfield & R. A. Sternbach (Eds.), *Handbook of psychophysiology.* New York: Holt, 1972.

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

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

Janisse, M. P. Pupil size and affect: A critical review of the literature since 1960. *Canadian Psychologist,* 1973, *14,* 311-329.

Kahneman, D. *Attention and effort.* Englewood Cliffs, New Jersey: Prentice-Hall, 1973.

Kahneman, D., & Beatty, J. Pupil diameter and load on memory. *Science,* 1966, *145,* 1583-1585.

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

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.

Kahneman, D., Tursky, B., Shapiro, D., & Crider, A. Pupillary, heart rate and skin resistance changes during a mental task. *Journal of Experimental Psychology,* 1969, *79,* 164-167.

Liberman, A. E. Some new data on the pupillary component in man. In L. G. Veronin, A. N. Leontiev, A. R. Luria, E. N. Sokolov, & O. S. Vinogradova (Eds.), *Orienting reflex and exploratory behavior.* Washington D.C.: American Institute of Biological Sciences, 1965.

Loewenfeld, I. E. Comment on Hess's findings. *Survey of Opthalmology,* 1966, *11,* 291-294.

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

Moos, R. H., & Engel, B. T. Psychophysiological reactions in hypertensive and arthritic patients. *Journal of Psychosomatic Research,* 1962, *6,* 227-241.

Nunnally, J. C., Knott, P. D., Duchnowski, A., & Parker, R. Pupillary response as a general measure of activation. *Perception and Psychophysics,* 1967, *2,* 149-155.

Paivio, A., & Simpson, H. M. The effect of word abstractness and pleasantness on pupil size during an imagery task. *Psychonomic Science,* 1966, *5,* 55-56.

Polt, J. M. Effect of threat of shock on pupillary response in a problem-solving situation. *Perceptual and Motor Skills,* 1970, *31,* 587-593.

Simpson, H. M., & Molloy, F. M. Effects of audience anxiety on pupil size. *Psychophysiology,* 1971, *8,* 491-496.

Stanners, R. F., & Headley, D. B. pupil size and instructional set in recognition and recall. *Psychophysiology,* 1972, *9,* 505-511.

Stanners, R. F., Headley, D. B., & Clark, W. R. The pupillary response to sentences: Influences of listening set and deep structure. *Journal of Verbal Learning and Verbal Behavior,* 1972, *11,* 257-263.

Vacchiano, R. B., Strauss, P. S., Ryan, S., & Hochman, L. Pupillary response to value linked words. *Perceptual and Motor Skills,* 1968, *27,* 207-210.

Wolpe, J., & Lazarus, A. A. *Behavior therapy techniques.* New York: Pergamon, 1966.

Woodmansee, J. J. Methodological problems in pupillographic experiments. *Proceedings of of the 74th Annual Convention of the American Psychological Association,* 1966, 133-134.

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