Extroversion, time of day, vigilance performance, and physiological arousal: Failure to replicate traditional findings*

ANTHONY GALE†, RAY BULL, VALERIE PENFOLD
MICHAEL COLES††, and ROSEMARY BARRACLOUGH
University of Exeter, Devon, Exeter, England

Three experiments are reported. Performance in the first may be used to predict performance in the second and third. The first examined the effect of extroversion, neuroticism, sex, and time of day on performance on a group-administered hour-long auditory vigilance task (N = 160; times of day: 0800, 1000, 1400, and 2000 h). Performance was related to none of these variables, but the task did yield a significant decrement in performance during the first 12 min (p < .05), and level of performance was well within the range typically reported for such tasks. Good and poor vigilance performers were then compared on two further separate tasks: (1) tonic and phasic EDA during habituation to repeated auditory stimulation (N = 22), and (2) tonic occipital EEG during alternated eyes-open and eyes-closed trials (N = 16). Of the EDA measures, tonic EDA (basal resistance) discriminated between good and poor vigilance performers. The resistance level of the good group increased during habituation trials (p < .01) and was also higher than that of the poor group following the ninth of 20 trials (p < .05). The tonic EEG of the good group was higher in amplitude than that of the low group (p < .05). These results are contrary to predictions made on the basis of arousal interpretations of individual differences in vigilance performance, since both the EDA and EEG measures indicated that good vigilance performers are lower aroused than poor vigilance performers.

Mackworth (1969) and Davies & Tune (1970) provide extensive reviews of the literature on vigilance performance. A number of consistent results have emerged concerning individual differences and certain environmental variables. We are concerned with the following: (1) introverts are better at vigilance tasks than extroverts, (2) there is diurnal variation in performance such that afternoon testing yields inferior performance to morning and early evening testing, and (3) time of day and personality effects may interact such that introverts are most efficient in the morning and extroverts most efficient in the afternoon. These summary statements are based on a number of experimental studies and are voiced in several reviews. Such data have been accounted for largely by interpretation of variation in vigilance performance in terms of arousal theory. At the outset of the present series of experiments, the robustness of these three findings was assumed. Our concern is mostly with physiological correlates of vigilance performance (e.g., Gale, Haslum, &

Penfold, 1971). Experiment 1 was designed merely as a large-scale screening procedure to provide criterion groups of poor (and extroverted) vigilance performers and good (and introverted) vigilance performers for further physiological study. Having selected such groups, we proposed to study physiological indices of arousal under vigilance task conditions. Employing a factorial design, we tested also for time of day and sex effects. We failed to obtain any individual difference in performance related to any of these independent variables; nevertheless, there were considerable individual differences in vigilance performance. Thus, Experiments 2 and 3 were designed to compare extreme performers on two simple physiological indices of arousal. Following interpretations of previous studies of individual differences in performance it was predicted that good vigilance performers would be characteristically more highly aroused physiologically (EDA and EEG) than would poor performers.

EXPERIMENT 1
EXTROVERSION, NEUROTICISM, TIME OF DAY, SEX, AND VIGILANCE PERFORMANCE

The Task

This was a 60-min auditory task consisting of a continuous sequence of digits presented at the constant rate of 1 digit/sec. There were 50 wanted

signals (ws) consisting of *three consecutive odd but unequal* digits. The ws/non-ws ratio was approximately 0.014. To permit a temporal analysis, the task was subdivided into five 12-min segments, with 10 ws per segment. The ws were prepared from random numbers, 1 to 9, and inserted within a body of random numbers (constrained in not having ws sequences). The ws were placed within each 12-min segment according to the following intervals, which were randomized for each segment: 48, 107, 16, 128, 73, 58, 30, 70, 11, 96, 63 sec, with an interval between ws in adjacent segments not exceeding 128 sec. The S was instructed to place a tick for *every individual* digit in the task on squared arithmetic paper but to *respond to ws by recording the final digit of the ws sequence*. This was, therefore, a continuous response task with differential response for ws. The task was preceded by a practice session, which included a 5-min practice under true task conditions. The instructions, the practice session, and the task proper were recorded on tape and administered binaurally through headphones.

Physical Arrangements

The experiment was conducted in a 12-booth language laboratory. The booths were modified to exclude vision, by covering the forward face with black card and extending the partitions to the side such that no S could see any other S. The headphones through which the task was administered served also to eliminate auditory stimulation by other Ss. The individual tape decks were covered with a raised black platform, on which the S's response sheet was pinned. The only objects the S could see were the headphone cable, the pencil, and the response sheet. The task was presented to all booth headphones simultaneously from the central console.

Procedure

The Ss completed the E.P.I. (Form B). These were collected, as were the Ss' watches. The instructions and practice task were played. The instructions included a break for queries concerning the task. After the practice session, the practice sheets were collected. The task then began. Ss attended at one of four times: 0740, 1000, 1400, or 1940 h. Total administration of instructions and task took 90 min. Ss were tested in groups ranging in size from 8 to 12 persons per session.

Subjects

The Ss were 190 undergraduate and postgraduate students at the University of Exeter, age range 18 to 32 years. Of the original sample of

Table 1
Experiment 1: Mean Detection Scores for the Subgroups (Maximum Possible = 50)

Time of Day	Extraverts (EPI Score > 14)				Introverts (EPI Score ≤ 14)				Total N =
	M	N	F	N	M	N	F	N	
0800	34.1	12	32.2	8	38.6	10	32.1	10	40
1000	33.1	10	34.5	10	38.6	10	30.9	10	40
1400	34.7	10	37.3	10	35.8	10	36.9	10	40
2000	39.4	11	33.5	9	34.4	16	34.0	4	40
Total N =		43		37		46		34	160

190, 8 Ss were eliminated (1 with language difficulty, 1 falling asleep, 2 failing to understand instructions, 4 refusals to complete task). This left 182 Ss. Their vigilance response sheets were scored prior to scoring of the personality questionnaires.

Results

(1) Product-moment correlations were computed for extroversion:neuroticism (r = −0.04), extroversion: detections (r = −0.02), and neuroticism:detections (r = −0.01). None of these correlations were significantly different from zero. Four separate correlations, between extroversion and performance were computed for the times of day. None of these reached significance (0800 h, r = −0.23; 1000 h, r = +0.12; 1400 h, r = −0.13; 2000 h, r = +0.12).

(2) Analyses of variance: A cutoff criterion was required to enable equality of cell membership. Since previous work had focused on extroversion rather than neuroticism, the 22 Ss with neuroticism scores equal to or greater than 17 were then removed from the sample, yielding 160 Ss. An analysis of variance on detections was computed on the data shown in Table 1, with main effects being time of day and extroversion and sex being collapsed. Ss with extroversion scores of 14 or more were classified as extrovert. None of the F ratios for main effects or interactions were significant. The analysis was then repeated, subdividing total scores for the five 12-min segments. The main effects in this analysis were extroversion, time of day, and time in task. Only time in task yielded a significant F ratio (F = 3.97, p < .01). Mean values for the five segments are plotted for all 160 Ss in Fig. 1. The difference between the first and second segment (first 12 min against second 12 min) is significant beyond the 5% level (t = 2.5, two-tailed). Other comparisons between segments failed to reach significance.

(3) A series of t tests were computed to test for sex differences, either overall or for times of day taken separately. None of these reached significance.

(4) Visual inspection of the data for male extroverts only indicates the possibility of a time of day effect.

However, none of the t tests employed to test this possibility reached significance.

(5) Commission errors were few, ranging between a within-time-of-day mean of 1.9 and 3.7 per group; overall mean, 3.04. There were no significant differences for personality, sex, or time of day (t tests). Nor was there any relationship between detections and commission errors (r = +0.08).

EXPERIMENT 2
VIGILANCE PERFORMANCE AS A PREDICTOR OF ELECTRODERMAL ACTIVITY

Subjects

Two groups were selected (randomizing for time of day in Experiment 1) on the basis of vigilance performance; a good group (6 males, 5 females; range of detections, 46-49) and a poor group (5 males, 6 females; range, 10-21). There were thus 22 Ss, with 11/group. The groups were not different in regard to extroversion or neuroticism scores. Group membership was coded so that the E conducted the experiment blind. Classification of Ss was revealed to him only when scoring of the records was completed.

The Task

After a 4-min rest period, 20 tones (sine wave, 1,000 Hz, 80 dB) of 5 sec duration were presented through a loudspeaker on a variable schedule. The interstimulus interval varied between 45 and 80 sec in 5-sec steps. The schedule was stored on a tape programmer, and administration was automatic. Total task time was 20 min. The S was instructed to relax and was informed that a uniform sound would be presented at irregular intervals; he was not to count the sounds or try to anticipate them; he was to keep his eyes shut but not fall asleep.

Physical Arrangements

The S lay on a reclining hairdresser's chair in a soundproof cubicle. A light gray curtain was suspended from ceiling to floor on all walls. A low level of illumination (8.5 fL) was provided by a lamp suspended above and behind the S's head.

Apparatus

Silver-silver-chloride electrodes were placed on the volar surface of the upper segment of the index and middle finger of the nonpreferred hand (left in all cases). Surface area of the electrode was 1 cm², and current passed by the GSR meter was 10 microA. The output of the meter was fed to a dc amplifier and written out on a San'ei polygraph at a paper speed of 10 mm/sec, maximum sensitivity being 300 ohms. Stimulus onset was marked automatically on the record on a separate channel.

Scoring the EDA

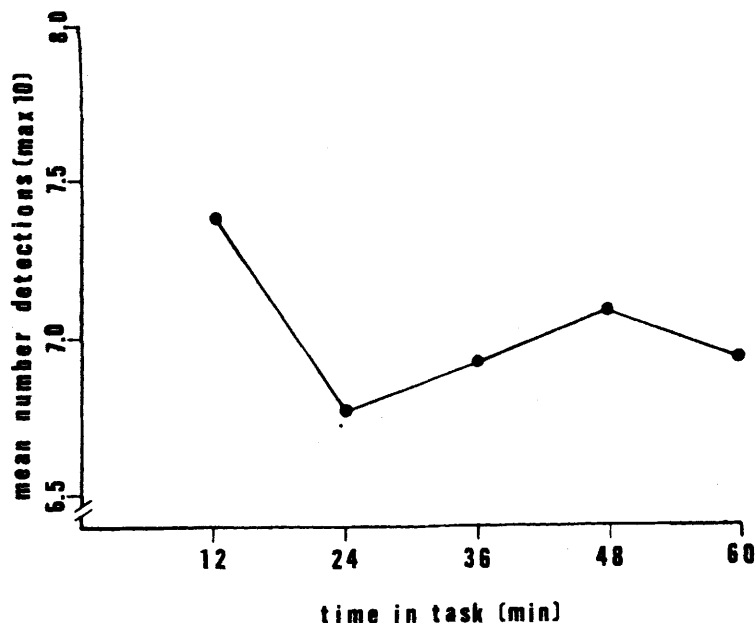The following measures were taken:

Fig. 1. Experiment 1: The vigilance decrement (N = 160).

(1) tonic base level prior to each stimulus, (2) latency of response to the first stimulus, (3) total number of responses, (4) number of responses before first, second (consecutive), and third (consecutive) no response, (5) magnitude of response to first stimulus (change in log conductance), and (6) number of spontaneous responses. Any deflection within 10 sec of the onset of the stimulus was classed as a response. On the few occasions when responses occurred within this 10 sec but following two consecutive no responses, they were classed as spontaneous.

Results

(1) Comparison of good and poor vigilance performers: Only the comparison of tonic levels yielded significant results. First, the good group increased resistance from the second to final trial, 10 Ss in the good group having higher resistance values prior to Trial 20 than prior to the first trial, with 1 S tying (sign test, p < .02). Secondly, after Trial 9, the good group's tonic level was always higher than that of the poor group (separate t tests, two-tailed; for each trial, p < .05 on all occasions). See Fig. 2.

(2) Correlations between measures: For all Ss, the following measures intercorrelated (Spearman rank correlations were employed in this case because of the noncontinuity of the sample): total responses: total spontaneous responses (p < .01); total
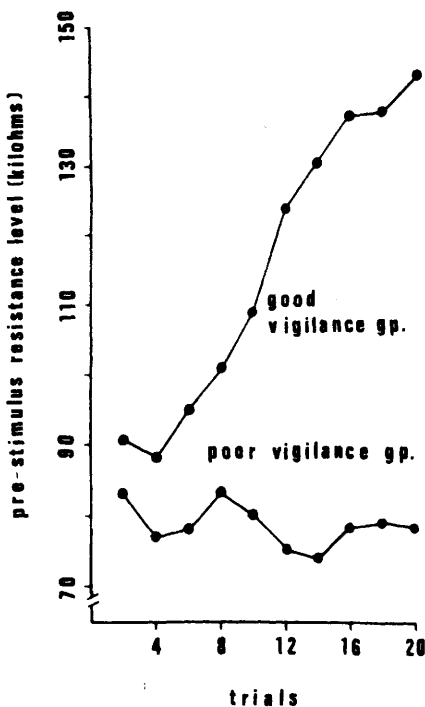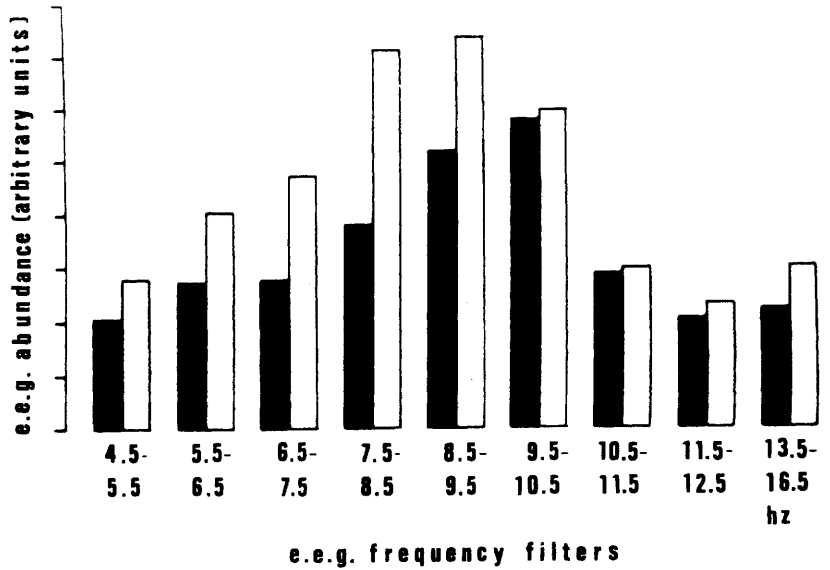
Fig. 3. Experiment 3: Resting EEG for good (white) and poor vigilance performers. (Scores for eyes open and eyes closed are collapsed.)

responses: latency to first response (p < .01); total responses to three consecutive no responses (habituation): latency to first response (p < .05); and habituation: spontaneous responses (p < .02) (all two-tailed tests).

EXPERIMENT 3
VIGILANCE PERFORMANCE
AS A PREDICTOR OF
RESTING EEG

Subjects

Only 16 Ss were now available from the population of Experiment 2. The good vigilance group contained three males and five females, and the low vigilance group, four males and four females. Classification of Ss was unknown to the E and was revealed only when scoring was complete.

The Task

The S was instructed to open and shut his eyes on command. There were 10 2-min trials, eyes open and eyes closed alternating. Total task time was 20 min.

Physical Arrangements
and Procedure

The physical arrangements were as for Experiment 2. Instructions to open and close eyes were given over an intercom.

Apparatus

Electrode placement, EEG recording, and automated analysis were identical to that employed by Gale, Coles, & Blaydon (1969). In this case, however, EEG abundance values for the different frequencies were converted by a Lion Systems Development A-D converter and stored on punch tape.

Results

For each S, there were five trials of eyes open and five trials of eyes closed

for each of the frequency values shown in Fig. 3. Each trial consisted of 24 5-sec epochs. Alternate epochs were sampled, yielding a 60-sec sample per trial. Analyses of variance were computed for each filter taken separately, the main effects being open/closed, vigilance performance, and time in task. For all filters, the difference between eyes open and eyes closed was significant beyond the 5% level, eyes closed yielding higher scores. For three filters, the difference between good and poor vigilance performers was significant beyond the 5% level (4.5-5.5 Hz: F = 5.10; 6.5-7.5 Hz: F = 5.17; 7.5-8.5 Hz: F = 5.55). In all cases, since this was a significant main effect for vigilance performance, the difference holds for eyes open and eyes closed scores taken together (see Fig. 3). None of the main effects for time in task were significant. However, there was a significant triple interaction for 5.5-6.5 Hz. A further analysis of variance on the eyes-closed data taken alone revealed that the poor vigilance performers decreased in amplitude, in this frequency, over time in task (F = 4.82, p < .05). This reduction was distributed equally over the task as a gradual trend and could not be attributed to a difference between pairs of trials.

DISCUSSION

The results may be summarized as follows: (1) In the vigilance task employed, extroversion, neuroticism, time of day, and sex appear to have no effect on performance; (2) no differences are found for extroversion, either for total number of detections or for changes in patterning of performance during the task; (3) time

Fig. 2. Experiment 2: Tonic resistance during habituation trials (N = 22).

of day and extroversion do not interact in a systematic manner; (4) the task yields a significant decrement in performance over the first of five 12-min segments; (5) good vigilance performers are not different from poor performers on response measures of EDA; (6) however, good performers show a significant increase in tonic resistance during the latter half of habituation trials; (7) correlations between the various measures of EDA response are consistent with previous findings; and, finally, (8) when eyes-closed and eyes-open trials are collapsed, and resting EEG of good performers is of greater abundance than that of poor performers.

Thus, we have failed to replicate previous work on time of day and personality. In addition, we have demonstrated that on two physiological indices of arousal, good vigilance performers appear to be lower aroused than do poor vigilance performers. This finding runs counter to interpretations of the individual differences in performance obtained in earlier work. We shall discuss the present experiments in turn.

Experiment 1

Are there features of our task which might help to explain the failure to replicate earlier findings? Group administration is a likely candidate, since this has been shown to reverse the normal relationship between extroversion and performance (Colquhoun & Corcoran, 1964). There are two difficulties here. Firstly, what exactly is a "group"? Our Ss were in the same room, but all sources of interaction were eliminated. Zajonc (1965) claims that the presence of others is arousing; indeed, Gale, Lucas, Nissim, & Harpham (in press) have demonstrated the influence of social interaction on the EEG. Thus, group membership might raise extrovert scores and lower introvert scores, yielding a null difference. However, in the present case, it can only be mere *knowledge* of group membership which influences performance, since the Ss were virtually in isolation. Secondly, if in fact performance is influenced by this variable, how does it interact with other variables to produce the invariance obtained? A very complex explanation would be called for. This problem must be acknowledged by any critic of our design for, whatever the defect may be, it leads to equality of scores on the dependent variable for all groups.

The task itself involves continuous responding. Since this would produce excessive generation of reactive inhibition for extroverts (e.g., Eysenck, 1967), one might expect the task to *accentuate* the differences for extroversion. Indeed, this is what we hoped for when we designed the experiment. Continuous response is a novel feature. However, many explanations of the course of vigilance performance, particularly those employing habituation as a central concept, do depend upon a notion of continuous response, albeit central and/or physiological rather than behavioral (e.g., Mackworth, 1969). More recent work has shown that each individual digit in this task has a differential effect on physiological variables. Thus, there are moment-to-moment shifts in arousal as a function of the sequential cue value of individual digits, i.e., consecutive odd numbers induce a progressive accumulation of arousal in both EEG and skin conductance (Gale et al, 1971; Gale, Bull, & Haslum, in press; Gale, Haslum, & Lucas, in press). Such findings corroborate arousal theory accounts of fluctuations of arousal in the vigilance task and confirm the validity of the present paradigm.

Again, one might argue that the wanted signal is not of the type typically employed in vigilance tasks, where the emphasis is on detection of perceptual change rather than on sequences demanding cognitive evaluation of signals. However, the task is a modified version of the original Bakan studies on individual differences and performance (Bakan, 1959; Bakan, Belton, & Toth, 1963). The Colquhoun (1960) study did employ a more conventional brand of task. But *explanations* of individual differences in performance have focused on vigilance tasks *in general* and do not discriminate between the different variants. Results from earlier laboratory studies have been extended to a variety of work and performance situations. It is, of course, questionable whether the Colquhoun and Corcoran study, which presents the model for the interaction of time of day, isolation, and personality effects, itself satisfies minimal requirements for a vigilance or inspection task. There, task time was a mere 15 min. There must be something exceptionally different about our task to disqualify it as a vigilance task proper. The times of day employed, size of sample, and the criteria for classification of Ss are in accord with earlier designs, and the decrement and overall level of performance are characteristic of earlier data. Even though there are no differences for extroversion or time of day, the task does induce systematic variation in performance which allows for discrimination between high and poor groups in the physiological measures of Experiments 2 and 3. Thus, some unidentified variable which yields systematic physiological variation accounts in part for variation in performance. An analysis of variance design is ideally suited to test for the main and interaction effects and is indeed the only one which would allow for the inferences drawn in previous work. We have included product-moment correlations because this statistic was employed by Colquhoun and Corcoran (1964); again, these provide no support for previous work. We can only conclude that Experiment 1 was an acceptable attempt to replicate earlier findings which failed. Regrettably, negative findings must be accounted for more rigorously than corroborative data. However, when performance on this task is employed as an independent variable in Experiments 2 and 3, more puzzling data emerge.

Experiment 2

Our prediction here was that poor vigilance performers would, in physiological terms, be lower aroused than good vigilance performers. This prediction may be justified on a number of grounds. Activation or arousal theories of vigilance suggest that high arousal is associated with detections and low arousal with omissions. Davies and Tune (1970) cite several studies showing association of physiological indices of low arousal during low detection rate periods. There is, however, little evidence to show discrete covariation of these indices with individual detections or omissions, though Gale et al (1971) have shown that EEG abundance decreases as the likelihood of a signal increases and then increases following detections. Secondly, extroverts are characterized as poor vigilance performers and also as chronically lower aroused; thus, there is EEG evidence indicating higher EEG abundance for extroverts under certain conditions (Gale, et al, 1969). Thirdly, we have ourselves demonstrated a relationship between physiological reactivity and individual differences in detection rates. In one study (Coles & Gale, 1971), EDA reactivity (habituation, latency of response to first stimulus, total number of responses) and number of detections in a separate vigilance task were directly related. Gale and Lynn[1] have shown EDA and EKG reactivity to be directly related to detections and inversely related to a measure of "stimulus hunger." The results of the present experiment are completely at variance with those findings, and, although there were procedural differences in those studies, we can think of no ready explanation. However, there is a possibility of inconsistency in any study which predicts both high resting levels and

high response levels for the same individual(s). According to Wilder's (e.g., 1957) law of initial value, there can only be responsiveness of great magnitude when resting levels are low. If the law does hold, then there is inconsistency in arousal explanations of individual differences in vigilance performance. If, for example, introverts are already highly aroused, then presumably they have a lower potential for response. (This problem is neglected by Mackworth and by Davies and Tune.) Thus, it might be argued that our good groups, in showing a reduction in tonic resistance, were capable of greater response magnitude. However, there is a tendency, though nonsignificant, for the *poor* group to be more responsive. As for the task itself, the correlations between the different measures of reactivity indicate that it is a proper measure of EDA habituation; these correlations are generally reported in the literature (e.g., Koepke & Pribram, 1966; Bull & Gale, 1971). There is, finally, a paradox in terms of the time when increase in resistance level occurs (see Fig. 2). We would have expected this to occur as a possible accompaniment to vigilance decrement in the poor group. However, the good group shows an increase, while the poor group maintains a stable level. The failure of a reduction in conductance in a simple habituation task is, in itself, unusual, and groups yielding such data are worthy of further study.

Experiment 3

Here again, we predicted that good vigilance performers would be characterized by a low-amplitude "aroused" EEG. The findings are both contrary to our predictions and *consistent* with those in Experiment 2. Gale et al (1969) have shown this particular procedure to be capable of discriminating extroverts from introverts. Since no measure of responsiveness was taken in this case, the results refer only to tonic level. The finding of a group association of unidirectionality of arousal on two physiological indices measured on different occasions is again unusual and points to some unspecified but constant personality trait.

CONCLUSION

We have attempted to replicate earlier studies and to obtain further physiological support for the underpinning of arousal theory interpretations of those studies. We have failed to replicate earlier work, but cannot account for this failure. The physiological data, which indicates the presence of a constant source of individual variation in attentional tasks, runs counter to arousal theory explanations of individual differences in performance.

REFERENCES

BAKAN, P. Extraversion-introversion and improvement in an auditory vigilance task. British Journal of Psychology, 1959, 50, 325-332.

BAKAN, P., BELTON, J. A., & TOTH, J. C. Extraversion-introversion and decrement in an auditory vigilance task. In D. N. Buckner and J. J. McGrath (Eds.), *Vigilance: A symposium.* New York: McGraw-Hill, 1963. Pp. 22-33.

BULL, R., & GALE, A. The relationships between some measures of the galvanic skin response. Psychonomic Science, 1971, 25, 293-294.

COLES, M. G. H., & GALE, A. Physiological reactivity as a predictor of performance in a vigilance task. Psychophysiology, 1971, 8, 594-599.

COLQUHOUN, W. P. Temperament, inspection efficiency, and time of day. Ergonomics, 1960, 3, 377-378.

COLQUHOUN, W. P., & CORCORAN, D. W. J. The effects of time of day and social isolation on the relationship between temperament and performance. British Journal of Social & Clinical Psychology, 1964, 3, 226-231.

DAVIES, D. R., & TUNE, G. S. *Human vigilance performance.* London: Staples Press, 1970.

EYSENCK, H. J. *The biological basis of personality.* Springfield, Ill: Thomas, 1967.

GALE, A., BULL, R., & HASLUM, M. Cumulative expectancy, subjective report of alertness and electrodermal activity. Psychophysiology, in press.

GALE, A., COLES, M., & BLAYDON, J. Extraversion-introversion and the EEG. British Journal of Psychology, 1969, 60, 209-223.

GALE, A., HASLUM, M., & LUCAS, B. Arousal value of the stimulus and EEG abundance in an auditory vigilance task. British Journal of Psychology, in press.

GALE, A., HASLUM, M., & PENFOLD, V. EEG correlates of cumulative expectancy and subjective estimates of alertness in a vigilance-type task. Quarterly Journal of Experimental Psychology, 1971, 23, 245-254.

GALE, A., LUCAS, B., NISSIM, R., & HARPHAM, B. Some EEG correlates of face-to-face contact. British Journal of Social & Clinical Psychology, in press.

KOEPKE, J. E., & PRIBRAM, K. H. Habituation of GSR as a function of stimulus duration and spontaneous activity. Journal of Comparative & Physiological Psychology, 1966, 61, 442-448.

MACKWORTH, J. F. *Vigilance and habituation: A neuropsychological approach.* Harmondsworth: Penguin Books, 1969.

WILDER, J. The law of initial value in neurology and psychiatry. Facts and problems. Journal of Nervous & Mental Diseases, 1957, 125, 73-86.

ZAJONC, R. B. Social facilitation. Science, 1965, 149, 269-274.

NOTE

1. Gale, A., & Lynn, R. Vigilance performance, physiological reactivity and stimulus hunger. In preparation.