Quarterly Journal of Experimental Psychology
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/pqje19

The breakdown of vigilance durning prolonged visual search

N. H. Mackworth [a]

[a] Medical Research Council, Applied Psychology Research Unit, Cambridge
Version of record first published: 07 Apr 2008.

To cite this article: N. H. Mackworth (1948): The breakdown of vigilance durning prolonged visual search, Quarterly Journal of Experimental Psychology, 1:1, 6-21

To link to this article: http://dx.doi.org/10.1080/17470214808416738

These problems should be amenable to experimental investigation, and for this purpose the use of patients being treated with electric convulsion therapy seems to be specially suitable, and these have already been used to a limited extent for this purpose. I think one can assume that the electric shock paralyses all cerebral activity so that the clinical features of concussion are very closely reproduced.

In conclusion, I fear I have wandered from the subject of our discussion, and do not feel I have done justice to it. I hope, however, to have said enough to make you feel that the traumatic disorders of remembering are worthy of consideration.

REFERENCES

1. FERRIER, D. (1878). *The Localisation of Cerebral Disease* (London).
2. RUSSELL, W. R. and NATHAN, P. W. (1946). *Brain*, **69**, 280.
3. RUSSELL, W. R. (1948). *Edin. Med. Jnl.*, **55**, 92–99.

(*Manuscript received 7th October, 1947.*)

THE BREAKDOWN OF VIGILANCE DURING PROLONGED VISUAL SEARCH[1]

BY

N. H. MACKWORTH

(*From the Medical Research Council Applied Psychology Research Unit, Cambridge*)

I

INTRODUCTION

1. *The General Problem.* The deterioration in human performance resulting from adverse working conditions has naturally been one of the most widely studied of all psychological problems. Amongst other possibilities, the stress arising from an unusual environment may be due either to physico-chemical abnormalities in the surroundings or to an undue prolongation of the task itself. This paper is concerned with the latter form of stress, as it has been found to occur in one particular type of visual situation; a later publication will more fully discuss the implications of these and other visual and auditory experiments (Mackworth, 1948).

The relevant literature on the deterioration in perceptual efficiency resulting from prolonged work makes depressing reading; not only is it scanty, but it is also rather contradictory. For example, in 1890, William James could confidently define the nature of attention—but nearly 50 years later both Woodworth (1938) and Bills (1934) were dubious about the whole concept of attention, the latter maintaining that the term had lost its meaning from an identification with the conscious results of the process rather than with the process itself. Head (1926) used the term vigilance to describe both a physiological and a psychological readiness to react, and the present writer also believes that vigilance is a useful word to adopt, particularly in describing a psychological readiness to perceive and respond, a process which, unlike attention, need not necessarily be consciously experienced.

[1] This paper has been based on one given to the first meeting of the Experimental Psychology Group.

Thorndike held that for tasks up to two hours in length "the work grows much less satisfying or much more unbearable, but not much less effective." Robinson (1934) criticised this view because it was based on experiments on accuracy of work at tasks which gave the subjects immediate and reliable information on any trend in their working efficiency. This was an interesting point because Wyatt and Langdon (1932) studied performance trends in work which had no such accuracy self-check and here they did find deterioration in accuracy of industrial inspection when the workers had been 30–45 minutes at the task; this decrease continued for about 90 minutes and was followed by an irregular recovery towards the end of the four-hour working spell. Ditchburn (1943), on the other hand, found that look-out duties in a practical form began to deteriorate almost as soon as the work was started and vigilance fell to a minimum level within only a few minutes. Anderson *et al.* (1944) undertook some interesting and detailed experiments with synthetic radar equipment which suggested that "in general, to ensure the most efficient performance, it does not appear wise to prolong daily operating periods more than 40 minutes, if such periods of operation are to be repeated daily without intervening days of rest. On the other hand, it is possible that occasional operating periods of as much as 4 hours in duration may be tolerated without marked reduction in efficiency."

2. *The Specific Problem.* The Second World War provided many situations which involved prolonged visual search. In 1943, for example, the Royal Air Force had to determine the optimum length of watch for airborne radar operators on anti-submarine patrol. It was suspected that working efficiency was deteriorating due to overlong spells at the radar screens. The situation was, in fact, thought to resemble one described by Shakespeare:

> "For now they are oppress'd with travel, they
> Will not, nor cannot, use such vigilance
> As when they are fresh."—(*The Tempest*, Act 3, Scene 3.)

Common sense, however, had given no precise answer to the problem because discussion with operational personnel showed that in practice their working spells at the radar sets over the Atlantic were anything from one-half to 2 hours in length (Craik and Mackworth, 1943). Laboratory studies were therefore started, initially for this wholly practical reason, to determine over what length of time accuracy could be maintained in work of this sort. It was arbitrarily decided that the synthetic laboratory situation need attempt to reproduce only the more general features of watch-keeping duties. It therefore seemed necessary to provide a series of visual signals which were all difficult to perceive because the subject had no more than a glimpse of each of these barely visible stimuli. It was also thought important to ensure that the signals were presented only occasionally, so that the lengthy searching nearly always drew a blank, although many rather similar signals of no importance at all were constantly being encountered and having to be disregarded. Isolation of the watcher seemed a significant factor in the situation, and care was taken to ensure that the subject had no reliable objective yardstick by which he could gauge his own performance.

II

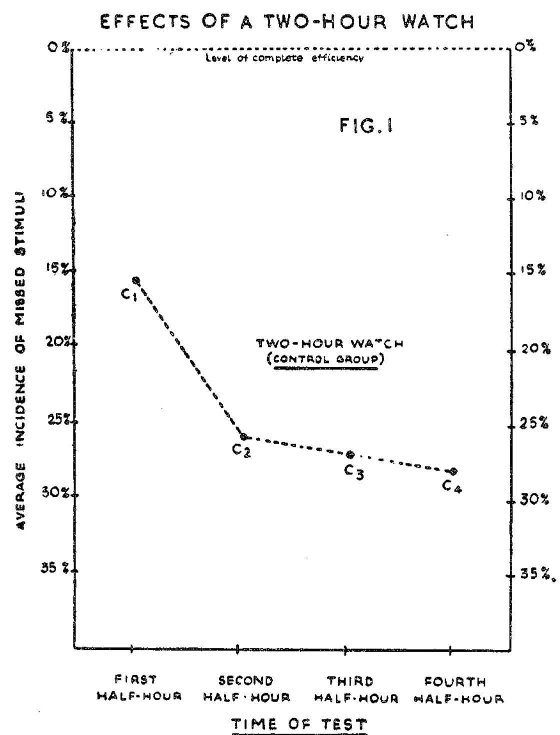
Method, Procedure and Subjects

1. *General Procedure.* The Clock test was devised to produce a visual situation into which these general psychological factors could be introduced. The subject sat in a wooden cabin in a room entirely by himself and looked at a black pointer 6 inches in length. This normally moved on in steps (like the seconds hand of a large clock)

in front of a plain white vertical surface. It jerked on to a new position once every second, 100 of these movements making the full circle. There were no scale markings or reference points of any kind on this white background.

While the subject was looking at this ordinary movement, he was told that "every now and again, at long and irregular intervals, instead of the pointer moving like so, it will move through double the usual distance. This double movement gives the effect of one unusually large movement forward. Press the response key as soon as you notice one of these double-length movements. Though I want you always to press the key as quickly as possible, do not think it is ever too late to press it, if you suddenly remember having seen the signal."

The pointer was mounted in such a way that its tip traced out a circle 10 inches in diameter. Every ordinary movement of the pointer therefore moved the tip through 0·3 inch and every double movement pushed it on 0·6 inch. Since the subject

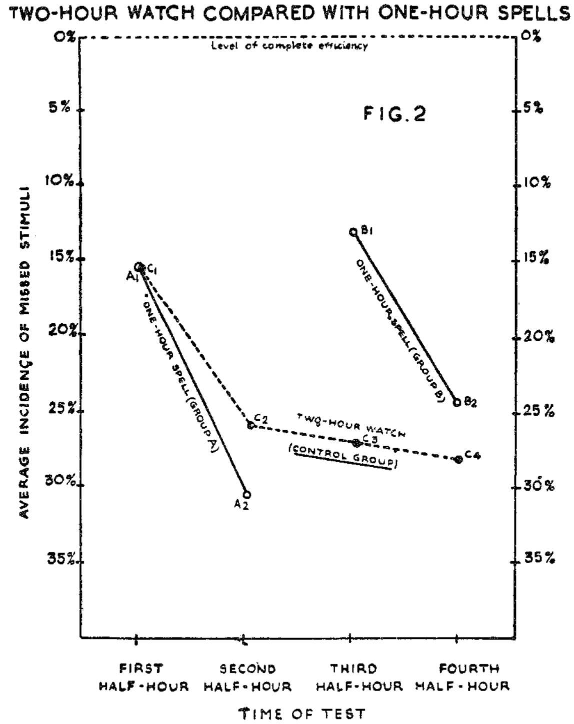


was 7 feet away, this 0·3 inch difference between the two displacements subtended an angle of 12 minutes at the eye, and the subject was picking out 24-minute movements from the usual "background" of 12-minute movements. The width of the pointer itself subtended an angle of 8 minutes at the tip and 16 minutes at the broad end.

The subject had a preliminary run on the first five double-length stimuli of the series. During this practice spell the experimenter watched the clock with the subject and immediately after the first one or two double-length movements he said "Now!" or "That's one!" to make sure the subject had understood the task. In this preliminary period the experimenter told the subject whether his responses were right or wrong.

The subject then went on to do the actual experiment after being told that it would last for 2 hours. In this test the double-length movements of the hand occurred only 12 times in 20 minutes, at intervals of $\frac{3}{4}$ minutes, $\frac{3}{4}$, $1\frac{1}{2}$, 2, 2, 1, 5, 1, 1, 2 and 3 minutes, in that order. Ten further minutes of the ordinary uneventful movement completed the first half-hour of the experiment. The second, third and fourth half-hours of the test were identical with the first, and followed straight on without any obvious break in the presentation. In case the subject realised this repetition he was asked to lend the experimenter his wrist watch for the duration of the test. As far as possible all extraneous noises were excluded from the room; the apparatus made a constant droning noise and also some other faint irregular sounds to mask all auditory cues to these double-length movements of the clock hand. The

TWO-HOUR WATCH COMPARED WITH ONE-HOUR SPELLS



scoring part of the equipment was in an adjacent room and this recorded the time taken by the subject to press a morse key in response to each of the 48 stimuli given during the 2-hour experiment. If no response came from the subject within 8 seconds of the stimulus being applied, it was assumed that he had missed that particular signal. The experimenter gave the subject no knowledge of results at all during the actual test or at the end of the experiment.

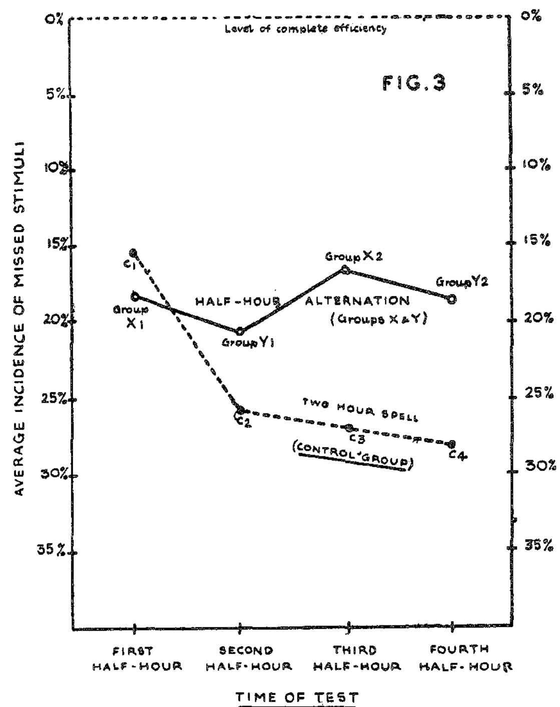
2. *Specific Procedures.* The Clock test in this basic form has been given under a range of different conditions:

(*a*) *Two-hour Watches.* 25 R.A.F. cadets were each tested once for a two-hour spell, and their results formed the scores shown in Figure 1.

(*b*) *One-hour Watches.* 50 other R.A.F. cadets were also tested, one at a time, for one hour each. They reported in pairs; one man did the first hour of the test, and then the second man (who had been waiting outside the room) took over for the second hour. The 25 people who did the first hour were Group A in Figure 2; the 25 men who did the second hour were Group B. A comparison could therefore be made between the average performance of 50 men divided into two groups each working for one hour with that of the control group who worked for a two-hour spell.

(*c*) *Half-hour Watches.* 50 further R.A.F. cadets were then tested in the following way. They reported in pairs—Subject X and Subject Y; Subject X did the first and third half-hours of the test and Subject Y did the second and fourth half-hours. This meant that results were obtained from the half-hour alternation of two groups (Group X and Group Y of Figure 3). During their half-hour break these subjects had no opportunity of determining their level of accuracy on their initial spell.

TWO-HOUR WATCH COMPARED WITH HALF-HOUR SPELLS

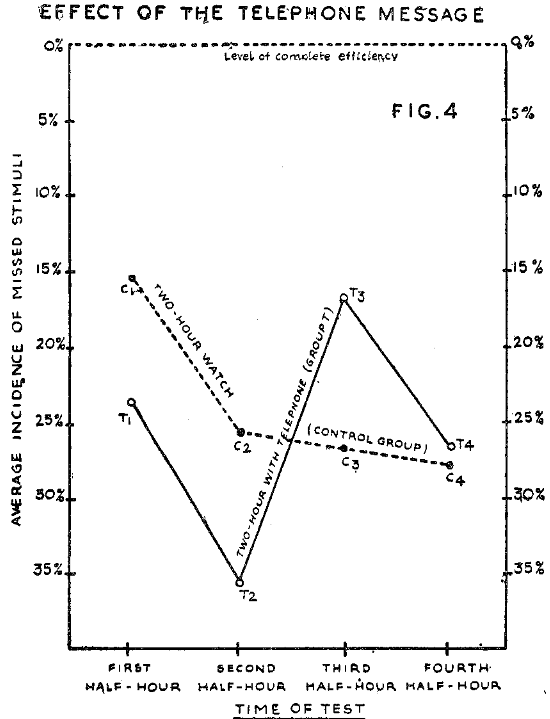


(*d*) *Two-hour Watches with the Telephone Message.* 25 other R.A.F. cadets were then tested exactly as for the two-hour control group experiments, but one alteration was made to the conditions. The subject was given a hand microphone telephone and told he would receive various instructions over this from time to time during the test. In fact, only one standard message was given, and this always arrived in the middle of the ten minutes of negative stimuli ending the second half-hour of the test. The message therefore immediately preceded the beginning of the third half-hour, the results for which are plotted as point T_3 in Figure 4. The exact technique was that a warning buzzer sounded and as soon as the subject answered, he was asked over the telephone "to do even better for the rest of the test; to try to see even more of these double-length movements of the pointer."

(*e*) *The Briefing Experiment.* An investigation was also made of the effects of briefing the subjects. In this, the six subjects did a two-hour run exactly as the test for the two-hour control group, but again with one alteration in the conditions.

These men were given an extra display board to watch as well as the Clock hand. The whole of their briefing was given before the test began and each subject was told that the board represented the position of his imaginary aircraft in relation to the area of sea being patrolled. During the two-hour experiment, a plotting arrow crept very slowly across the map, so that the subject had to glance at this display only occasionally to determine his supposed position. He was asked to concentrate on the Clock

EFFECT OF THE TELEPHONE MESSAGE



pointer and told to be ready at all times for the double-length movements of the hand, and warned to be especially on the alert when the indicator arrow was passing through the "danger area" marked on the map, as more signals were to be expected during that period. The plotting arrow was timed to enter this special area just before the third half-hour of the test, and left it at the end of the third half-hour.

III

Results

1. *Results of Two-hour Watches* (*Control Group*). It is clear from Figure 1 that the men began to miss more signals after they had been working for half-an-hour at the task.

Mean Percentage Incidence of Missed Signals

First Half-hour (C_1)	*Second Half-hour* (C_2)	*Third Half-hour* (C_3)	*Fourth Half-hour* (C_4)
15·7% (46/293)	25·8% (75/291)	26·8% (79/295)	28·0% (82/293)

This reduction in working efficiency after half-an-hour was not likely to have been due to uncontrolled factors in the experiment, since the odds were at least 19 to 1 against the difference between C_1 and C_2 being due to chance; the difference between the two means was exactly three times its standard error, when twice would have shown a significant difference. The differences between C_1 and C_3 and between C_1 and C_4 were also statistically significant since the critical ratios here were 3·3 and 3·5 respectively. On the other hand, the slight fall in efficiency between C_2 and C_4 was not statistically definite since the critical ratio here was only 0·6.

2. *Results of One-hour Watches* (*Group A and Group B*). Here again there was a marked difference between the performance of the subjects during their first half-hour and their second half-hour.

MEAN PERCENTAGE INCIDENCE OF MISSED SIGNALS

Nature of Group	*First Half-hour* (A_1)	*Second Half-hour* (A_2)	*Third Half-hour* (B_1)	*Fourth Half-hour* (B_2)
Group A ..	15·3% (45/294)	30·4% (91/296)	—	—
Group B ..	—	—	13·0% (38/292)	24·2% (72/297)

The differences between A_1 and A_2 and between B_1 and B_2 were statistically sound since critical ratios of 4·4 and 3·7 were obtained. On the other hand there was no significant difference between A_1 and C_1 since here the critical ratio was 0·1—nor was there any real difference between A_2 and C_2, as the critical ratio was 1·2.

It will be noted that bringing in the fresh subjects (Group B) on the third half-hour of the test reduced the incidence of missed signals to one-half of the value obtained from the control group. This difference between C_3 (26·8% missed) and B_1 (13·0% missed) was statistically significant as the critical ratio was found to be 4·2. But this advantage again lasted for only half-an-hour, because during the last half-hour of the test there was no real difference between C_4 and B_2, as the critical ratio here was only 1·1.

3. *Results of Half-hour Watches* (*Groups X and Y*). Statistical analysis showed that under the conditions of rapid, half-hour alternation there was no significant difference between the average performance on the first half-hour of the test and that during any other period of the experiment. In Figure 3 the point X_1 was not statistically different from either Y_1 or X_2 or Y_2, the critical ratios being 0·8, 0·5 and 0·1 respectively.

MEAN PERCENTAGE INCIDENCE OF MISSED SIGNALS

Nature of Group	*First Half-hour* (X_1)	*Second Half-hour* (Y_1)	*Third Half-hour* (X_2)	*Fourth Half-hour* (Y_2)
Group X ..	18·1% (53/293)	—	16·5% (51/289)	—
Group Y ..	—	20·7% (61/294)	—	18·5% (54/291)

Comparison with the two-hour control group in Figure 3 shows that on the first half-hour of the test the initial points X_1 and C_1 of the two curves were not significantly different (the critical ratio being 0·8) when both groups were fresh to the work. But during the last hour-and-a-half of the experiment there was a highly significant difference between the two-hour group and the half-hour alternating groups. With the points C_2, C_3 and C_4 taken together and the points Y_1, X_2 and Y_2 combined, the critical ratio was 4·1. Comparing these points separately there were significant differences between C_2 and X_2 and between C_4 and Y_2—critical ratios being 3·0 and 2·7, but the difference between C_2 and Y_1 was not significant, the critical ratio being 1·5. This was probably due to chance factors producing a slightly less efficient group of people for Group Y. (It might have been thought that because Group Y had to wait before doing the test, this would have had some deleterious effect on point Y_1 which would not have affected point X_1. This seems unlikely since in the previous experiment there was no significant difference between point A_1 and B_1 of Figure 2).

Though there was no statistically sound difference in Figure 3 between X_1 and Y_1 or between C_1 and Y_1, the increased incidence of errors at C_2 over X_1 was quite reliable, the critical ratio being 2·2. The apparent improvement between points X_1 and X_2 and between Y_1 and Y_2 was not confirmed on statistical analysis as the critical ratios came to 0·5 and 0·7 respectively.

4. *Results of Two-hour Watches with the Telephone Message* (*Group T*). Analysis of the average incidence of missed signals (see Figure 4) suggested that listening for the telephone at the same time as looking for the double movements of the clock hand was a definite handicap; because during the first hour of the test (before the message was given) there was a significant difference between the two-hour control and the two-hour telephone groups. The critical ratio for the points C_1 and T_1 was 2·4, and for the points C_2 and T_2 was 2·5. But the telephone message dramatically reduced the number of missed signals; it raised the average level of efficiency for the third half-hour to a standard which was usually obtained by fresh subjects, not by men like these who had been working for more than an hour at the task.

TOTAL INCIDENCE OF MISSED SIGNALS

Nature of Group	*Time after Telephone Message*			
	Within 13 minutes (Stimuli 1–7 of 3rd half-hour)	*Within 18–25 minutes (Stimuli 8–12 of 3rd half-hour)*	*Within 35–43 minutes (Stimuli 1–7 of 4th half-hour)*	*Within 48–55 minutes (Stimuli 8–12 of 4th half-hour)*
Telephone message group . .	28	22	45	33
Two-hour control group . .	49	30	48	34
Ratio . .	0·57	0·73	0·94	0·97

The difference between C_3 and T_3 in Figure 4 showed a significant advantage in favour of the telephone group, as the critical ratio was 3·0. The point T_3 was not significantly different from C_1 (the critical ratio being 0·3), but T_3 was just significantly better than T_1 (the critical ratio being 2·0). This improvement did not last for more than half an hour since the difference between C_4 and T_4 was not statistically definite, and here the critical ratio was 0·4. A subjective improvement following the telephone message was reported by the majority of the subjects and this also lasted for about half an hour.

A further analysis of the data was made to ensure that the improvement in performance did in fact extend over all the stimuli in the third half-hour, because it was just possible that the effect had been confined to the signals given immediately afterwards. The total number of signals missed by the 25 men was calculated for the first seven signals given after the message and for the remaining five signals given in the third half-hour. The data from the fourth half-hour were subdivided in the same way—and the results from the two-hour control group provided a comparison. Since the telephone message was given five minutes before the beginning of the third half-hour, the evidence was obtained at the time intervals after the message. (See p. 13.)

In other words there was no doubt that the introduction of the telephone message produced an effect which lasted for 25 minutes, but this disappeared about 35 minutes after the application of the stimulus.

5. *Results of the Briefing Experiment.* Unlike the experiment with the telephone message, this briefing technique had no effect at all on efficiency during the third half-hour of the experiment, or at any other time. The following table compares the average incidence of missed signals obtained from this briefed group of subjects with that made by the two-hour control group. There was no significant difference between the two sets of scores.

MEAN PERCENTAGE INCIDENCE OF MISSED SIGNALS

Nature of Group	*First Half-hour*	*Second Half-hour*	*Third Half-hour*	*Fourth Half-hour*
Briefed Group ..	18·1% (13/72)	26·4% (19/72)	27·8% (20/72)	27·8% (20/72)
Two-hour Control Group	15·7% (46/293)	25·8% (75/291)	26·8% (79/295)	28·0% (82/293)

6. *Results of a Stimulus-by-Stimulus Analysis of the Performance Trend.* An attempt was made to detect any fluctuations in working efficiency due to the spacing of the signals in the test series, i.e., the pattern made in time. Two difficulties arose which necessitated the use of average results taken from a large number of subjects in searching for these minute-by-minute variations in accuracy. The two disturbing factors were:—(1) each man had some entirely random fluctuations in his level of vigilance; and (2) all stimuli were not of entirely uniform difficulty because a double-length movement was probably easier to detect when the clock hand was pointing in certain directions than in others. This was important because a double-length movement could start at any one of the hundred possible positions of the clock hand; the spatial position of such test signals was therefore entirely randomised within the circle traced out by the tip of the clock hand.

For the *first and second half-hours* of the experiment, results were available from 214 subjects, i.e., the average incidence of missed signals was calculable on these subjects for each of the 24 stimuli taken in turn. Not enough data were available to do this for the *third and fourth half-hours* of the experiment, but there the two time patterns could be superimposed, and, for example, the evidence from the first, second, third and fourth, etc., signals in the third half-hour was taken together with that from the first, second, third and fourth, etc., signals respectively of the fourth half-hour. This meant there were 278 presentations of each stimulus in this combined third half-hour/fourth half-hour series.

The results are plotted in Figure 5; they confirmed the marked difference in overall accuracy between the first half-hour and the rest of the test. There was also some suggestion of a warming-up effect within the first five minutes of the experiment. This was followed by a downward trend in efficiency during the remainder of the first half-hour of the test. During the second and subsequent half-hours this falling-off was arrested, and performance oscillated around the same low level of accuracy. That the bulk of the experiment was being done by the subjects in an inaccurate manner is clear when one sees from Figure 5 that the subjects *usually* made twice or even three times as many mistakes compared with their optimum performance. Three to five minutes after the beginning of the experiment the men averaged 8% or 9% errors, but for the last three half-hours of the test, the same people gave average error scores which fluctuated between 19% and 28% missed signals.

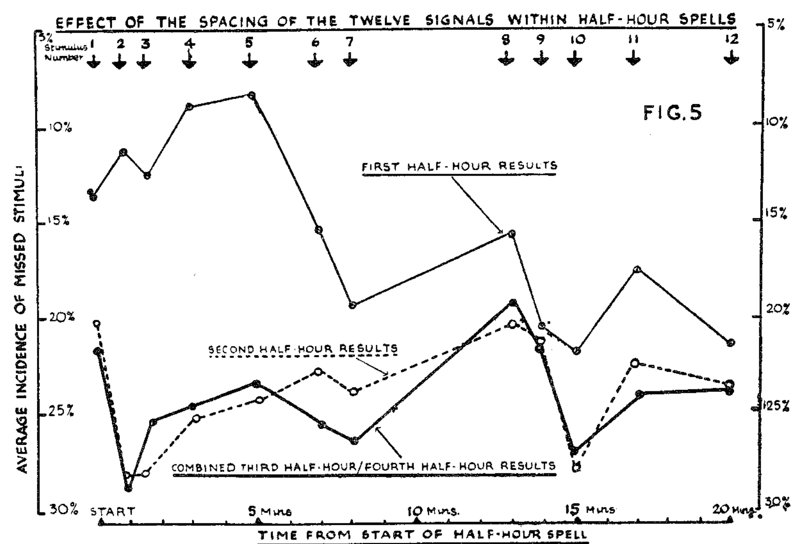


Figure 5 emphasises that apart from the long-term trend, there were similarities between the stimulus-to-stimulus trends within the various half-hour spells. At first glance, there seems no good reason why, for example, the seventh stimulus in each of the half-hour series should, on the average, always have proved more difficult than the eighth. It is believed that the similarity between the stimulus-by-stimulus records for each half-hour is explicable only on the grounds that the spacing of the signals in time drew a definite pattern of performance by producing minute-by-minute changes in attitude which arose from the apparent occurrence or non-occurrence of the signals. If one compares the three curves shown in Figure 5, it seems clear that (apart from the initial warming-up effect during the first five minutes) the signals that were most difficult to see were those presented in relatively rapid succession at intervals of $\frac{3}{4}$—1 minute apart, i.e., the second and third stimuli, and the ninth and tenth stimuli in the series. This may have been due to the fact that the instructions and the test, as a whole, suggested that a good deal of time elapsed between signals.

The subjects, therefore, lowered their vigilance after noticing the first or eighth signals—particularly as these followed, and were therefore in direct contrast with, blank spells lasting ten or five minutes, which were entirely free from meaningful activity. Once a signal has been seen, the subject may miss the next stimulus if he does not expect the next signal to appear for some little time, and it does in fact appear before he expected it.

An attempt has been made to analyse the data statistically to see whether any reliable evidence could be obtained about this effect of the length of the blank spell preceding the stimulus on accuracy of work. The data from the first half-hour of the experiment have been excluded owing to the marked general trend previously described and the study has, therefore, been confined to the last hour-and-a-half of the experiment. The results were classified into three main groups according to the length of the preceding inter-stimulus time interval.

EFFECT OF LENGTH OF PRECEDING INTER-STIMULUS TIME INTERVAL ON MEAN PERCENTAGE INCIDENCE OF MISSED SIGNALS

Length of preceding blank spell ..	$\frac{3}{4}$–1 *minute*	1$\frac{1}{2}$–3 *minutes*	5–10 *minutes*
Average incidence of missed signals	25·9% (637/2460)	24·1% (594/2460	20·4% (201/984)

The critical ratio between the 25·9% reading and the 20·4% reading came to 3·5, but the critical ratio between the 25·9% incidence and the 24·1% incidence was only 1·5; the critical ratio between the 24·1% score and the 20·4% score was also only 1·3. The subjects were, therefore, definitely worse on signals given after blank spells of one minute or less than they were with those presented after an inter-stimulus interval of 5 to 10 minutes.

7. *Results of a Study of the Individual Differences.* In considering the differences between people in their ability to do this kind of work, the index of performance selected was the number of missed signals per half-hour. In view of the sharp

FREQUENCY DISTRIBUTION TABLE FOR MISSED SIGNALS

Number of signals missed (out of 12)	*Fresh Group*		*Tired Group*	
	%	*Number*	%	*Number*
0	24·9	(56)	14·7	(33)
1	24·9	(56)	13·8	(31)
2	16·9	(38)	15·1	(34)
3	11·1	(25)	12·9	(29)
4	9·3	(21)	8·8	(20)
5	3·1	(7)	13·8	(31)
6	4·4	(10)	8·0	(18)
7	4·0	(9)	8·4	(19)
8	0·9	(2)	2·2	(5)
9	0·5	(1)	1·8	(4)
10	—	—	0·5	(1)
11	—	—	—	—
12	—	—	—	—
	Total ..	255 readings	Total ..	225 readings

deterioration in efficiency found after half an hour at the work, it seemed reasonable to divide the results into two categories—fresh and tired subjects.

(i) *Fresh Subjects:* Results from people who were doing their first half-hour on the test and from those starting again after half an hour's break.

(ii) *Tired Subjects:* Results from men who had been working more than half an hour on the test.

A study of the frequency distribution table (see p. 16) showed the wide range of ability amongst either fresh or tired subjects. Secondly, although the distributions obtained from the fresh and tired groups overlapped, the tired group tended to have a greater proportion of readings in the lower levels of efficiency. For example, it will be seen that about half the readings in the fresh group were in the categories where either all the signals were seen or only one stimulus was missed; only about a quarter of the records from the tired men reached this high pitch of efficiency. These marked individual differences are not related either to visual acuity grades or group intelligence test scores on verbal and spatial relationship material.

IV

Discussion

1. *Introductory Points.* The above experiments suggest that lapses in visual perception become more frequent (i) when the Clock test has been proceeding for more than half-an-hour, (ii) when the subject has recently responded to one visual signal and is not expecting another since he imagines the stimuli are few and far between—and (iii) when the subject is expecting an auditory stimulus while doing the visual task. It is perhaps permissible to mention briefly here that other studies by the writer have also indicated that (iv) lack of sleep is liable to lead to more missed signals on this test and (v) so also to a lesser extent will an increase of the atmospheric temperature of the working environment to a reading about that of the normal body temperature.

Conversely, the conditions that are known to improve performance include (i) a short rest of half an hour and (ii) the receipt of a telephone message. Similarly, fewer signals pass unnoticed (iii) if while they do the test the subjects are allowed to have a progress record giving them immediate knowledge of their results, or (iv) if they take 10 mgms. of amphetamine sulphate by mouth one hour before the start of the test. Either of these last two procedures keeps performance at the initial high level of accuracy throughout the two-hour spell.

Various possible interpretations of these facts can be attempted but none of the explanations appears wholly satisfactory to the writer. It is at least clear from the various psychological influences that can be used to alter the achievement level that the phenomenon is central in origin and is not due to changes in the peripheral visual mechanisms. A relevant point is that Carpenter (1948) has shown that, although the average blink rate on the Clock test does give a rising trend similar to the half-hour to half-hour incidence of missed signals, this is a coincidental and not a causal relationship.

It might be thought that these researches were dealing only with alterations in the general attitude of the subjects towards the whole test situation—some fading of interest in external events, perhaps leading to a condition like the lowered standard phenomenon described by Bartlett (1941) where the subjects on a highly skilled task became less self-critical and allowed progressively larger and larger errors to occur before they regarded these as calling for action. This view of the Clock test results is very hard to maintain owing to the absence of effect with the briefing technique; presumably on this theory this ought to have given some performance improvement by modifying the general attitude of the objects during the part of the test spell that they had been told was particularly important. Similarly there was no "suggestion effect" improvement in performance when men did the test after taking some unspecified tablets which in fact were pharmacologically inert. If the subjects were simply becoming bored with the work, one feels that they ought also to have been *more*, and not less, alert immediately after the occurrence of some slightly unusual event in their environment such as one of these long signals. It would still be necessary to know what caused this fading of interest and why it was not usually progressively more marked with each half-hour at the work. On this last point Oldfield has suggested that the flattening of the work curve (which was also found by Anderson *et al.* (1944)), might be taken to indicate that the normal and relatively stable state was that represented by the large error score in the last hour-and-a-half of the test spell and that the first half-hour results were due to an abnormal and unstable level of alertness.

2. *Discussion of the Main Downward Trend.* The writer considers that the least unsatisfactory way of interpreting the findings on the Clock test is to regard the condition under consideration primarily as an example of a state of inhibition—in the sense of the decrease or absence of a response which was the result of some form of positive stimulation. Since this decrease was apparently due to the repeated presentation of stimuli incorporated in the test situation itself, it would seem reasonable to regard the lowered efficiency as an example of internal inhibition. Other interpretations of the facts are possible and all are hard to bring to the experimental test.

In his paper on internal inhibition, Oldfield (1937) remarked that no attempt was being made to explain the intimate mechanics of all cortical activity; internal inhibition was regarded as a state which may descend on any one of a number of different cortical functions. The present writer has used the same approach; internal inhibition has been studied because it seems to have been found in circumstances which at first were being investigated entirely empirically to investigate a practical problem. The terminology of conditioned response theory has not been adopted with the unjustified assumption that the principles of the conditioned response provide a comprehensive solution of the mysteries of all cortical function.

In discussing the characteristics of situations leading to internal inhibition, Oldfield (1937) makes the uniformity of the total sensory field a first requirement; this uniformity is usually obtained in the form of simple, regular, discrete changes in the sensory environment, because after a time such regular changes tend to become accepted as uniformity. The Clock test satisfies this primary condition in view of the extreme regularity of the movements of the clock hand in *space*—round and round it goes, always taking the same short steps except for an occasional double-length signal. As regards regularity in *time*, the clock hand invariably moves with a regular rhythm which is not disturbed even by the double-length stimuli.

In his searching review of the relationship between conditioned response and conventional learning experiments, Hilgard (1937) made the point that the distinction between reinforcement and reward, sometimes permitted to separate conditioning from learning, is found to disappear when the conditioned response situation is carefully examined; Hilgard and Marquis (1940) have further developed this concept, including among their specific examples the learning experiments by Grindley (1932) and the studies on children by Ivanov-Smolensky (1927). Investigations of this sort have been grouped under the general heading of "Instrumental conditioning." The term instrumental conditioning is employed when the correctness or otherwise of the response determines whether a reward is given or withheld, and when the response can, therefore, be regarded as instrumental in obtaining the reward. The essential feature of instrumental conditioning is that the unconditioned stimulus leading to the original reaction is entirely different from the reinforcing stimulus which follows the conditioned response.

Food reward experiments satisfy a primary need of the organism, but it is clear that when the experimental subjects are adult human beings an effective reward need not necessarily satisfy a primary drive. Indeed a neutral stimulus may acquire the reinforcing properties of a reward if it has been repeatedly associated with a primary reward; Hilgard and Marquis (1940) term this "secondary reinforcement" or "derived reinforcement." In adult humans a very effective form of derived reinforcement is the token reward of knowledge of results, i.e., immediate information, supplied while the work is being done, about the success with which the prescribed task has been undertaken (Crawley, 1926; Grindley and Elwell, 1939; Mace, 1935).

When such investigations are borne in mind, it becomes clear that the Clock test situation readily lends itself to a description in conditioned response terms. During this account it has been found convenient to refer to the double-length movements of the Clock hand as the long signals, and to the ordinary single-length movements as the short signals.

The *original voluntary instructed response* was the movement of the arm and hand to press the reaction key.

The *unconditioned stimulus* was present only during the brief initial practice spell, and not at all during the main experiments. Even in the practice spell, it usually had to be given only once or twice in the first minute or so of the demonstration period. It took the form of the experimenter's remark, "Now!"—a shortened version of the implied instructions, "Press the key, now!" This comment was made immediately after a long signal, and it seems logical to regard it as the unconditioned stimulus, since by itself, without any other stimulus, it would have led to the original voluntary instructed response.

The *conditioned stimulus* was the long signal given by the Clock hand. This positive stimulus had to be differentiated from the mass of similar stimuli coming from the short signals made by the Clock hand.

The *conditioned voluntary response* occurred when the subject pressed the response key on detecting a long signal.

Reinforcement was normally given by the experimenter only during the practice period, and consisted of his comment "Yes, that was right" if the subject had responded correctly, *or* "No, that was wrong" if the key had been pressed after a short signal rather than a long one, *or* "You missed one there," if the subject had failed to notice a long signal. In other words the reinforcement was a derived reinforcement—the token reward of knowledge of successful results, and the token punishment of information on failures.

This description of the Clock test can be summarised as follows:—

Conditioned Stimulus	*Unconditioned Stimulus*	*Original Voluntary Response Conditioned Voluntary Response*	*Reinforcement*
Long movement of the clock hand	Command of "Now!"	Pressing response key	Comment of "Yes, that was right," etc. (Knowledge of results).

Very rapidly, during the practice spell, this situation gave an entirely new significance to what was, at first, the quite neutral stimulus of a long movement of the clock hand. This result is in keeping with the experience of previous workers who have found that the conditioned voluntary response is established quickly if it is formed at all.

Since the term internal inhibition demands a more precise definition it is necessary to attempt to define the nature of the condition more fully. It is likely that the exact form it took was that of experimental extinction, since this can usually be considered as the absence or weakening of a conditioned response upon repeated application of the conditioned stimulus without any reinforcing agent. Under the particular circumstances in question, this was taken to be the repeated presentation of the long signal without any provision of knowledge of results. Further evidence in support of this view that the long-term downward trend in performance from half-hour to half-hour was a manifestation of experimental extinction was found in the prevention of the usual increase in error score by the supply of knowledge of results (Mackworth, 1948).

It was interesting also to find that the sudden presentation of a novel distracting stimulus in the form of the telephone message immediately restored the partially extinguished response because this strongly suggested a disinhibitory effect, particularly as the briefing experiment which introduced no sudden unexpected change in the experimental situation failed to improve performance. But the length of time during which this disinhibition was effective was much longer than has usually been reported in Pavlovian experiments (Hull, 1934).

3. *Discussion of the Minor Fluctuations in Efficiency.* Although the experimenter did not usually give progress information to the subjects, and although this was apparently the main factor in the performance decline obtained in the Clock test, it remained possible that the subjects had a vague and uncertain source of knowledge of results in their own experience of the test situation. Figure 5 suggests that in fact this was so—that the subjects were collecting *some* information about their performance from the task itself as they went along. If they had not done so, then there would not have been any resemblance at all between the shape of the stimulus-to-stimulus curve for one half-hour compared with that obtained for another. This is believed to have been why the experimental extinction was only partial and not complete, i.e., that to some extent the influence of expectancy and self-instructions replaced the initial reinforcing stimulus of knowledge of

results. This resulted in second-order fluctuations being superimposed on the main trend of the working efficiency curve.

Gibson (1941) has emphasised in a valuable review the distinction between two forms of preparatory set—*expectancy* (expecting stimulus objects) and *intention* (intending to react). Mowrer, Rayman and Bliss (1940) have shown that expectancy can be varied independently from intention by demonstrating that subjects were definitely slower in reacting to a series of sounds when previous instructions had led them to expect either sounds or lights, than when they were told to expect only sounds. This divided attention effect may explain the deterioration in accuracy that was found during the first hour of the Clock test when the subjects had to listen for a telephone message while at the same time undertaking the ordinary visual task.

Similarly the finding by Mowrer (1940)—that a deterioration in auditory reaction time was more likely if the signal came in before it was expected, rather than after—would seem to be in keeping with the finding that errors on the Clock test were usually higher after the shorter inter-stimulus intervals.

V

Summary

Laboratory researches on prolonged visual search led to the development of an experimental situation in which the trend of ability at synthetic look-out duties was studied throughout long watchkeeping spells by the automatic production of occasional brief and barely visible signals. These stimuli were given at irregular time intervals; sometimes one or two minutes elapsing between signals, sometimes as long as ten minutes passing without significant incident. The results have been expressed in terms of the average error score for groups of healthy young men, but each subject was tested in a room entirely by himself the experimenter being in an adjoining room with the recording apparatus. These arrangements made it possible to keep the subjects largely unaware of their accuracies and inaccuracies at the task and so preserved an essential feature of vigilance testing.

Under these circumstances, efficiency at signal detection during a two-hour spell was definitely lower when the subjects had been watching the display for about half-an-hour. One-hour spells showed the same advantage in favour of the first half-hour compared with the second. But a half-hour rest following a half-hour watch was sufficient to allow a further half-hour spell to be undertaken straight away with no detrimental effect on accuracy. A sudden short telephone message in the middle of a two-hour watch produced a temporary improvement lasting only half an hour, but previous instructions that a particular period of the watch would be likely to contain rather more signals than usual had no demonstrable effect on performance.

Various possible interpretations of these changes in behaviour have been considered. The *main* downward trend in efficiency was eventually regarded as an example of the partial experimental extinction of a conditioned voluntary response, this extinction being the product of a lack of reinforcement arising from the absence of knowledge of results. The improvement following the distraction of the telephone message was interpreted as a form of disinhibition. There were also, however, *minor* stimulus-to-stimulus variations in efficiency which depended more on the general attitude of the subjects towards the test situation, particularly in regard to their opinion of the likelihood of a signal occurring at a given moment. For example, except in the first few minutes of the test the subjects were definitely less alert at detecting a signal when this was presented only one minute after the previous stimulus.

VI

Acknowledgments

Professor Sir Frederic Bartlett has greatly helped these researches with extensive advice, and helpful discussions on the problems have also been held with the late Dr. K. J. W. Craik, with Mr. G. C. Grindley and with Mr. R. C. Oldfield, but the interpretation is that of the writer. The statistical analysis was assisted by Mr. E. G. Chambers, and the experimental apparatus was constructed by Major H. C. L. Holden and Dr. E. Schuster. Mrs. Nixon de Weber gave most of the tests and prepared the illustrative diagrams. The Royal Air Force arranged the supply of the experimental subjects, and the writer is indebted to the Medical Research Council for financial support.

VII

References

1. Anderson, I. H. *et al.* (1944). Radar operator fatigue: the effects of length and repetition of operating periods on efficiency of performance. O.S.R.D. Report No. 3334, Research Report No. 6.
2. Bartlett, F. C. (1941). Fatigue following highly skilled work. 1941 Ferrier Lecture, published in 1943. *Proc. Roy. Soc. B.*, **131**, 247–257.
3. Bills, A. G. (1934). *General Experimental Psychology.* New York, London and Toronto.
4. Carpenter, A. (1948). The rate of blinking during prolonged visual search. *J. exp. Psych.* (*in the press*).
5. Craik, K. J. W. and Mackworth, N. H. (1943). Preliminary notes on the optimum length of ASV Mark III watch. (Unpublished report to Coastal Command, Royal Air Force).
6. Crawley, S. L. (1926). An experimental investigation of recovery from work. *Arch. Psychol.*, **13**, No. 85.
7. Ditchburn, R. W. (1943). Some factors affecting efficiency of work of look-outs. (Unpublished Admiralty Research Laboratory Report ARL/RI/84.46/o).
8. Gibson, J. J. (1941). A critical review of the concept of set in contemporary experimental psychology. *Psychol. Bull.*, **38**, 781–817.
9. Grindley, G. C. (1932). The formation of a simple habit in guinea pigs. *Brit. J. Psychol.*, **23**, 127–147.
10. Grindley, G. C. and Elwell, J. L. (1939). The effect of knowledge of results on learning and performance. *Brit. J. Psychol.*, **29**, 39–54.
11. Head, H. (1926). *Aphasia*, Vol. I. Cambridge.
12. Hilgard, E. R. (1937). The relationship between the conditioned response and conventional learning experiments. *Psychol. Bull.*, **34**, 61–102.
13. Hilgard, E. R. and Marquis, D. G. (1940). *Conditioning and Learning*, pp. 429. New York.
14. Hull, C. L. (1934). *Handbook of General Experimental Psychology*, ed. by Murchison. Chapter 9. The factor of the conditioned reflex. Clark University Press.
15. Ivanov-Smolensky, A. G. (1927). On the methods of examining the conditioned food reflexes in children and in mental disorders. *Brain*, **50**, 138–141.
16. James, W. (1890). *Principles of Psychology*, Vol. I, pp. 689. London.
17. Mace, C. A. (1935). Incentives. Medical Research Council, Industrial Health Research Board Report No. 72. H.M. Stationery Office, London.
18. Mackworth, N. H. (1944). Notes on the Clock test. Unpublished report to the Medical Research Council, A.P.U. 1, also Flying Personnel Research Committee Report No. 586.
19. Mackworth, N. H. (1948). Researches on the measurement of human performance. Medical Research Council Special Report Series. H.M. Stationery Office, London (*in the press*).
20. Mowrer, O. H. (1940). Preparatory set (expectancy)—some methods of measurement. *Psychol. Monogr.*, **52**, No. 233.
21. Mowrer, O. H., Rayman, N. N. and Bliss, E. L. (1940). Preparatory set (expectancy)—an experimental demonstration of its central locus. *J. exp. Psychol.*, **26**, 357–372.
22. Oldfield, R. C. (1937). Some recent experiments bearing on internal inhibition. *Brit. J. Psychol.*, **28**, 28–42.
23. Robinson, E. S. (1934). *Handbook of General Experimental Psychology*, ed. by Murchison. Chapter 12. Work of the integrated organism. Clark University Press.
24. Woodworth, R. S. (1938). *Experimental Psychology.* New York.
25. Wyatt, S. and Langdon, J. N. (1932). Inspection processes in industry. Medical Research Council, Industrial Health Research Board Report No. 63. H.M.S.O., London.

(*Manuscript received* 28*th December*, 1947.)